

UNIVERSIDAD DE DEUSTO
FACULTAD DE PSICOLOGÍA Y EDUCACIÓN
PROGRAMA DE INNOVACIÓN EDUCATIVA

COMPETENCIA EVALUATIVA DEL PROFESORADO
UNIVERSITARIO DE LOS GRADOS DE EDUCACIÓN INFANTIL Y
EDUCACIÓN PRIMARIA

AINHOA BILBAO MARTÍNEZ

Director

DR. AURELIO VILLA SÁNCHEZ

Bilbao, 2015

UNIVERSIDAD DE DEUSTO
FACULTAD DE PSICOLOGÍA Y EDUCACIÓN

PROGRAMA DE INNOVACIÓN EDUCATIVA



Tesis doctoral

COMPETENCIA EVALUATIVA DEL PROFESORADO
UNIVERSITARIO DE LOS GRADOS DE EDUCACIÓN INFANTIL Y
EDUCACIÓN PRIMARIA

Director

Dr. Aurelio Villa Sánchez

Doctoranda

Ainhoa Bilbao Martínez

Para Fran, mi compañero en la vida, en los sueños y en la pista de baile.

Por la enorme generosidad de hacer tuyas mis privaciones y desvelos.

*El reto que más me entusiasma en esta nueva etapa es poder devolverte todo el
tiempo y la atención que te he regateado estos años.*

Ojalá que mi “yo doctora” se parezca más a la compañera que te mereces.

C.W.M.

AGRADECIMIENTOS

La elaboración de una tesis doctoral es un proceso profundo e intenso que transforma a las personas. Este trabajo lleva marcas de todo lo que ha transcurrido en mi vida en los últimos años, de lo bueno y de lo malo, de los éxitos y de los fracasos. Ha sido escrito en épocas de calma y en días tristes de malas noticias, en escenarios amables y en otros que hubiera preferido no haber conocido. Cada capítulo tiene impresas parte de esas vivencias. En consecuencia, siento que esta tesis se lleva una parte importante de mi vida. Con ella, va todo lo que en este camino he tomado prestado de las personas que me rodean. Esos fragmentos de sus vidas, que he utilizado como fuente de estímulo y apoyo, también han quedado plasmados en este trabajo. Es justo, por tanto, que en estas últimas páginas que escribo para cerrarlo, les devuelva algo a cambio.

Desde los más profundo de mi ser, GRACIAS, ESKERRIK ASKO.

A *Aurelio Villa*, el patrón en esta aventura y en otras tantas. Junto a ti di mis primeros pasos en la docencia universitaria y en la formación permanente del profesorado. Mi estreno como investigadora tenía que ser así, de tu mano. Por todo lo que me has enseñado estos años y por lo que espero seguir aprendiendo.

A las *profesoras y profesores de la Universidad de Deusto y de Begoñako Andra Mari* que han permitido que me asome a su práctica evaluativa, un espacio sensible, tradicionalmente protegido de las miradas externas. También a las/os colegas que se tomaron un tiempo en orientarme y darme recomendaciones para encauzar la consulta de la mejor manera posible antes de iniciarla.

A *las/os estudiantes*, fuente inagotable de aprendizaje, que me han obsequiado con su tiempo y que me han sorprendido con la seriedad y la profundidad de sus aportaciones. Al alumnado de la Universidad de Deusto le agradezco especialmente la acogida y al de Begoñako Andra Mari la estimulación.

A *Jesús Marauri*, por su inestimable ayuda en este recorrido. Gracias por abrirme la puerta y dejar que me aproveche de tu experiencia previa.

A *Isabel Ariceta* por haber creído en mí y por traerme de vuelta a un proyecto que ya había dado por muerto.

A *Mireia Sanz*, por su extrema generosidad profesional y afectiva, porque es una suerte tenerte de jefa, de compañera y, sobre todo, de amiga.

A *Olga Cardeñoso*, por su disposición, su inspiración y su connatural *buenrollismo*. Espero que en esta nueva etapa busquemos oportunidades de seguir creciendo juntas.

A mi *gran familia*. A mis padres, Joseba y Maite, a mis hermanos Iker y Olatz, a mis tías/os y primas/os. Por vuestro apoyo incondicional. Por celebrar este producto casi más de lo que lo hago yo.

A *Josu, Elena, Eder y Oihana*, por todas esas noches de charla y pizza.

A mis *Fáciles* de mi vida y de mi corazón. A Pedro, Alber, Lorea, Pili, Maite, Manu y Josetxu. Por toda la salud mental que me han procurado esos buenos ratos con vosotros. Sois una fuente inagotable de alegría.

A todos aquellos que me han enseñado que *todo lo malo se va bailando*, porque desafortunadamente ha habido mucho que sacar y, afortunadamente, la receta ha funcionado.

A quien encarna a la vez todo lo anterior. A *Fran*, mi marido, compañero de vida y baile y, en los últimos meses, abnegado colaborador en la maquetación de este trabajo. Esto es para ti y gracias a ti.

ÍNDICE DE CONTENIDOS

PLANTEAMIENTO GENERAL DE LA INVESTIGACIÓN.....	31
CAPÍTULO INTRODUCTORIO	33
1. Justificación: relevancia, interés y pertinencia de la investigación	33
2. Objetivos de la investigación	42
3. Proceso de investigación.....	47
4. Estructura del trabajo	48
MARCO TEÓRICO	53
CAPÍTULO 1. LA EVALUACIÓN DEL APRENDIZAJE: COMPETENCIA DOCENTE CLAVE EN LA ENSEÑANZA UNIVERSITARIA	55
1. La docencia universitaria en el marco del EEES	56
1.1 El papel del profesorado en el proceso de convergencia.....	61
1.1.1 Espacios de actividad del profesorado en el panorama universitario actual: principales roles y funciones.	61
1.1.1.1 Principales modificaciones en el rol docente.	66
1.2 Principales dificultades en el desempeño del cometido que se atribuye al profesorado universitario en el panorama actual.	72
2. Perfiles profesionales del profesorado en el EEES	77
2.1 Caracterización general de los perfiles profesionales propuestos para el profesorado en la era de la convergencia.	79
2.2 Las competencias profesionales del profesorado universitario: perspectiva institucional de la Universidad de Deusto.	83
3. Evaluación del aprendizaje: competencia docente clave del profesorado universitario ...	88
3.1 Las constatadas disfuncionalidades de los sistemas tradicionales de evaluación del aprendizaje en la universidad.	93
3.2 El valor estratégico de la evaluación en la mejora de la calidad del aprendizaje universitario.	93

3.3 La evaluación del aprendizaje de las/os estudiantes como factor clave asociado a la calidad de la enseñanza universitaria.....	98
3.4 El papel de la evaluación en la consolidación del nuevo paradigma formativo de la educación superior.....	100
3.5 El riesgo de perpetuación de los modelos disfuncionales de evaluación del aprendizaje.....	101
3.6 Conclusión: investigación sobre la competencia evaluativa del profesorado en la formación básica de maestras/os y ciclo potencial de mejora.	102
4. El capítulo a grandes rasgos.....	105

CAPÍTULO 2. CONCEPTUALIZACIÓN: ESTABLECIMIENTO DE UN MARCO

COMÚN EN TERMINOLOGÍA EVALUATIVA..... 107

1. La evaluación del aprendizaje y su evolución histórica: distintas interpretaciones en la ciudad de Babel.....	108
1.1 Paradigmas de evaluación.....	113
1.1.1 La evaluación como indagación: revisión de conceptualizaciones.	116
1.2 Enfoques de evaluación.....	120
2. Construcción de una definición consensuada: ¿qué entendemos por evaluación?	123
3. Proceso de evaluación del aprendizaje.....	131
4. Funciones de la evaluación del aprendizaje	137
5. Tipología evaluativa. Revisión y ampliación de las clasificaciones clásicas en la evaluación del aprendizaje	143
6. Conclusión: la evaluación del aprendizaje de calidad.....	145
7. El capítulo a grandes rasgos.....	147

CAPÍTULO 3. LA EVALUACIÓN DEL APRENDIZAJE DE LAS/OS ESTUDIANTES

UNIVERSITARIAS/OS EN EL SISTEMA UNIVERSITARIO ESPAÑOL 149

1. Deficiencias generales de los sistemas tradicionales de evaluación del aprendizaje universitario en España.....	150
1.1 Descripción de síntomas.....	154
1.2 Resumen de las principales deficiencias.....	171
2. Efectos de los sistemas tradicionales de evaluación del aprendizaje en el alumnado y el profesorado universitario	172

2.1 Efectos de los sistemas disfuncionales de evaluación del aprendizaje en el alumnado universitario.....	172
2.2 Efectos de los sistemas disfuncionales de evaluación del aprendizaje en el profesorado universitario.	178
3. Factores asociados a las principales deficiencias en la evaluación del aprendizaje universitario y su persistencia	181
3.1 Condiciones o factores estructurales que originan disfuncionalidades evaluativas y/o favorecen su persistencia.	182
3.1.1 La influencia de las normas administrativas y reglamentos.	183
3.1.2 El incremento del número de estudiantes o la masificación de las aulas universitarias.	184
3.1.3 La creciente diversidad del alumnado universitario.	185
3.1.4 La incidencia de otras dificultades asociadas a la docencia en la evaluación del aprendizaje.	185
3.1.5 La sobrecarga docente y discente.....	186
3.1.6 Presiones sociales.....	186
3.1.7 Confusión o ambigüedad terminológica.	187
3.2 Condiciones o factores profundos que originan disfuncionalidades evaluativas y/o favorecen su persistencia.	187
3.2.1 La falta de cultura evaluativa en el sistema educativo en general y en el universitario, en particular.	187
3.2.2 El problema de la formación en evaluación.	189
3.2.3 Las creencias, teorías, concepciones o representaciones profundas sobre la evaluación y su resistencia al cambio.	190
3.2.3.1 Origen y características básicas de las representaciones evaluativas profundas.	193
3.2.3.2 El arraigo de las representaciones evaluativas profundas o su resistencia al cambio.....	194
4. Propuestas y sugerencias de mejora para la evaluación del aprendizaje universitario .	195
4.1 Iniciativas estructurales: sugerencias de mejora.	197
4.1.1 Revisión de las normas y condiciones administrativas que condicionan la evaluación del aprendizaje en la universidad.	198

4.1.2	Adopción de medidas para contrarrestar la sobrecarga que genera la masificación y la adopción de sistemas alternativos de evaluación.	198
4.1.3	Mejora de las condiciones laborales y profesionales generales en la docencia universitaria.	200
4.2	Iniciativas no estructurales o profundas: sugerencias de mejora.	200
4.2.1	Adopción de una actitud de humildad, revisión y autocrítica.	200
4.2.2	Formación en evaluación para el profesorado universitario.	201
4.2.2.1	Formación evaluativa cuantitativamente suficiente.	202
4.2.2.2	Formación evaluativa cualitativamente adecuada: la formación para el cambio de concepciones o representaciones profundas.	203
4.2.3	Desarrollo de experiencias de innovación en materia evaluativa.	206
4.2.4	Difusión de los avances, contribuciones y experiencias en materia evaluativa: investigación e intercambio.	207
4.2.5	Promoción y generalización de las experiencias de metaevaluación.	209
5.	Los sistemas tradicionales de evaluación del aprendizaje en la formación inicial de maestras/os de Educación Infantil y Primaria	212
6.	El capítulo a grandes rasgos.	216

CAPÍTULO 4. LA EVALUACIÓN DEL APRENDIZAJE EN LA UNIVERSIDAD DEL SIGLO XXI 221

1.	Referentes para la definición de una teoría de la evaluación del aprendizaje de las/os estudiantes en la universidad del siglo XXI.	222
1.1	Posición teórica de partida: la evaluación como indagación y como recurso estratégico para la promoción del aprendizaje.	223
1.2	La evaluación del aprendizaje en el contexto universitario de referencia.	229
1.2.1	Necesidades y exigencias del EEES con respecto a la evaluación del aprendizaje.	230
1.2.2	Referentes normativos para el desarrollo de la evaluación del aprendizaje en el contexto universitario.	233
1.2.2.1	Criterios y directrices europeas para la garantía de la calidad.	235
1.2.2.2	Orientaciones nacionales para la garantía de la calidad interna.	236
1.2.2.3	Estatuto del estudiante universitario.	238
1.2.3	La evaluación del aprendizaje en la formación básica de maestras/os.	239

1.2.3.1 La formación de maestras/os de Educación Infantil y Primaria en el contexto de las profesiones reguladas.	240
1.2.3.2 La evaluación del aprendizaje como competencia docente destacada y resultado de aprendizaje.....	241
1.3 Antecedentes en el establecimiento de criterios, principios o estándares para la evaluación del aprendizaje universitario.	244
1.3.1 Decálogos y manifiestos para la evaluación del aprendizaje universitario ...	246
1.3.2 Propuestas estructuradas de organismos y centros para la evaluación del aprendizaje universitario.....	254
1.3.3 Propuestas estandarizadas para la evaluación del aprendizaje	258
2. Conclusiones: un modelo de evaluación del aprendizaje hacia el que orientar la propuesta de competencias evaluativas específicas	262
2.1 Planificación estratégica del proceso de evaluación.	266
2.1.1 Potencialidad estratégica de la evaluación.	266
2.1.2 Evaluación alineada.	269
2.1.3 Selección metodológica justificada.....	270
2.1.4 Respuesta a la diversidad.	271
2.1.5 Evaluación de competencias.....	273
2.1.5.1 Evaluación globalizada.....	275
2.1.5.2 Evaluación continua y necesidad de seguimiento.	275
2.1.5.3 Evaluación basada en criterios y explícita.	276
2.1.5.4 Evaluación práctica y contextualizada.	276
2.1.5.5 Evaluación diversa y divergente.....	278
2.1.5.6 Evaluación participativa.....	279
2.1.6 Evaluación compartida o participativa.....	279
2.2 Comunicación y difusión de criterios y condiciones.	284
2.2.1 Explicitación y difusión de criterios y condiciones.....	285
2.2.2 Comprensión e interiorización de las condiciones básicas.....	286
2.3 Recogida y registro de evidencias evaluativas.....	288
2.3.1 Implementación ajustada.....	288
2.3.2 Recopilación y registro personalizado de evidencias de aprendizaje.....	289

2.4 Análisis de resultados.	290
2.4.1 Control de sesgos.....	290
2.4.2 Proceso de calificación.	293
2.5 Feedback y proacción.	294
2.5.1 Comunicación de los resultados de aprendizaje.	296
2.5.2 Retroinformación o feedback.....	296
2.5.3 Proacción para la mejora del aprendizaje.	302
2.5.4 Proacción o regulación de la enseñanza.	303
2.6 Revisión del proceso de evaluación: metaevaluación.	304
2.6.1 Revisión específica del dispositivo evaluativo.....	304
2.6.2 Proacción para la mejora de la evaluación.	308
2.7 Modelado en la formación básica de maestras/os de Educación Infantil o Primaria.	308
3. Modelo ITC	309
3.1 Subproceso 01: planificación consciente e integrada del proceso de evaluación. ...	310
3.2 Subproceso 02: comunicación y difusión de criterios y condiciones.	314
3.3 Subproceso 03: recopilación y registro de las evidencias de aprendizaje.	315
3.4 Subproceso 04: análisis y calificación de resultados.....	317
3.5 Subproceso 05: feedback y proacción.....	318
3.6 Subproceso 06: revisión del proceso de evaluación (metaevaluación).....	320
3.7 Subproceso 07: significación didáctico-pedagógica de la experiencia evaluativa. ...	322
4. El capítulo a grandes rasgos.....	323

CAPÍTULO 5. COMPETENCIA EVALUATIVA DEL PROFESORADO UNIVERSITARIO DE LOS GRADOS DE EDUCACIÓN INFANTIL Y PRIMARIA: PROPUESTA DE UN MODELO ESPECÍFICO PARA EL DIAGNÓSTICO Y LA FORMACIÓN 325

1. Competencia evaluativa del profesorado universitario dedicado a la formación básica de maestras/os: presentación del modelo.....	326
1.1 Evaluación a partir del modelo de competencia evaluativa propuesto.....	332
1.2 Formación a partir del modelo de competencia evaluativa propuesto.	335
2. Proceso de definición y ajuste.....	336

3. Fuentes consultadas en su definición	340
3.1 Antecedentes: análisis de otras propuestas que han delimitado la competencia evaluativa docente.	342
3.1.1 Competencia evaluativa: delimitación para el profesorado universitario de los grados vinculados a la formación básica de maestras/os de Educación Infantil y/o Primaria.	344
3.1.2 Competencia evaluativa: delimitación para el profesorado universitario.	345
3.1.3 Competencia evaluativa: delimitación para el profesorado.	345
3.2 Otras escalas o herramientas de probada validez que han atendido a la práctica evaluativa del profesorado en la formación universitaria.	348
3.3 Especificaciones sobre la competencia evaluativa incluidas en las escalas de evaluación de la calidad docente administradas en los grados vinculados a la formación básica de maestras/os.	351
3.3.1 Descripción del trabajo de campo.	352
3.3.2 Principales conclusiones.	354
4. Modelo de competencia evaluativa del profesorado universitario: desarrollo	360
4.1 Planificación estratégica de la evaluación para el aprendizaje.	362
4.2 Explicitación y transparencia.	367
4.3 Diversificación de la evaluación.	370
4.4 Implementación ajustada.	374
4.5 Interpretación y análisis de evidencias.	377
4.6 Regulación del aprendizaje y la enseñanza por medio de la evaluación.	381
4.7 Participación de las/os estudiantes en la evaluación.	386
4.8 Revisión del dispositivo evaluativo.	389
5. El capítulo a grandes rasgos	393

CAPÍTULO 6. PLANTEAMIENTO TEÓRICO DE LA COMPETENCIA EVALUATIVA DEL PROFESORADO: CONCLUSIONES DEL MARCO TEÓRICO 397

1. OBJETIVO A. Formular una propuesta revisada de las principales condiciones y características que debe reunir la evaluación del aprendizaje en el contexto de la formación básica de maestras/os	397

2. OBJETIVO B. Contribuir a la superación de la ambigüedad terminológica definiendo una clasificación revisada de tipologías o modalidades evaluativas que ofrezca un marco común de referencia.....	401
3. OBJETIVO C. Diseñar un modelo de competencia evaluativa para orientar procesos de formación y evaluación en el profesorado universitario dedicado a la formación básica de maestras/os	402
4. OBJETIVO D. Identificar los tópicos evaluativos presentes en las escalas de evaluación de la calidad docente administradas al alumnado en la formación básica de maestras/os (trabajo de campo intermedio)	403
MARCO EMPÍRICO.....	407
CAPÍTULO 7. JUSTIFICACIÓN DEL MARCO EMPÍRICO DE LA INVESTIGACIÓN	409
1. Diseño del marco empírico de la investigación.....	411
2. Población y muestra	415
2.1 Representatividad de la muestra.	417
2.2 Tamaño de la muestra.	421
3. Definición de variables	422
3.1 Variables dependientes.	424
3.2 Variables independientes.....	425
3.2.1 Variables demográficas.	426
3.2.2 Variables académico-institucionales.	426
3.2.3 Variables académico-profesionales.....	427
3.2.4 Posibles fuentes de influencia.	429
3.2.5 Variables extrañas.	431
4. Instrumentos para la recogida de la información	432
4.1 Proceso de creación de las herramientas.	436
4.2 Validez.	438
4.2.1 Análisis factorial exploratorio de los cuestionarios docente y discente para la determinación de la validez de estructura.....	440
4.2.1.1 Análisis factorial exploratorio de la subescala de importancia del cuestionario dirigido al profesorado.....	441

4.2.1.2 Análisis factorial exploratorio de la subescala de importancia del cuestionario dirigido al alumnado.	449
4.3 Fiabilidad.....	460
4.3.1 Consistencia interna del cuestionario del profesorado.	462
4.3.2 Consistencia interna del cuestionario del alumnado.	464
4.4 Posibilidades y limitaciones asociadas a las herramientas de recogida de información diseñadas.	466
5. Recogida de información.....	470
6. Previsión y desarrollo de los procedimientos de análisis.....	473
6.1 Análisis cuantitativos previstos para la consecución de los objetivos.	473
6.1.1 Entre la robustez y la cautela: el debate sobre la elección de procedimientos paramétricos y no paramétricos y la postura adoptada.	475
6.1.2 Elección de los procedimientos de análisis no paramétrico.....	478
6.2 Análisis cualitativos previstos para la consecución de los objetivos.	480
7. El capítulo a grandes rasgos.....	481
CAPÍTULO 8. PRESENTACIÓN DE RESULTADOS DEL MARCO EMPÍRICO DE LA INVESTIGACIÓN	483
1. Descripción de la muestra	484
2. Resultados relativos al Objetivo E. La competencia evaluativa docente desde la perspectiva del profesorado	487
2.1 Resultados Objetivo E1. Identificar los factores e indicadores de la evaluación del aprendizaje en los que mayor y menor dificultad expresa el profesorado, atendiendo al nivel de relevancia y realización que les atribuyen.....	488
2.2 Resultados Objetivo E2. Identificar relaciones entre distintas variables académicas, socio-demográficas o profesionales y la competencia evaluativa autopercebida.	493
2.2.1 Variables socio-demográficas y valoraciones docentes.....	496
2.2.2 Variables profesionales y valoraciones docentes.	498
2.2.3 Variables académico-institucionales y valoraciones docentes.	504
2.2.4 Autopercepción de competencia, formación evaluativa y valoraciones docentes.	508
2.3 Resultados Objetivo E3. Identificar las principales fortalezas, debilidades y dificultades para el ejercicio de la evaluación desde el punto de vista docente.	513

3. Resultados relativos al Objetivo F. La competencia evaluativa docente desde la perspectiva del alumnado.....	518
3.1 Resultados Objetivo F1. Identificar los factores e indicadores de la evaluación del aprendizaje en los que mayor y menor dificultad acusa el alumnado, atendiendo al nivel de relevancia y realización que les atribuyen.....	519
3.2 Resultados Objetivo F2. Identificar relaciones entre distintas variables académicas, profesionales o socio-demográficas y la competencia evaluativa percibida por el alumnado.	524
3.2.1 Variables socio-demográficas y profesionales.	529
3.2.2 Variables académico-institucionales.	535
3.2.3 Influencias en la valoración de la competencia evaluativa.	547
3.2.3.1 Fuentes de influencia potencial en las valoraciones del alumnado.	548
3.2.3.2 El impacto de la modalidad de respuesta en los resultados hallados.	556
3.3 Resultados Objetivo F3. Identificar las principales fortalezas y debilidades en la evaluación del aprendizaje desde el punto de vista discente.....	559
4. Resultados relativos al Objetivo G. Contraste de perspectivas sobre la competencia evaluativa docente.....	566
4.1 Valoraciones de importancia en los factores y dimensiones vinculadas a la competencia evaluativa docente.....	569
4.2 Valoraciones de realización percibida en los factores y dimensiones vinculadas a la competencia evaluativa docente.....	571
5. Resultados relativos al Objetivo H. Formación permanente en evaluación.....	575
5.1 Resultados Objetivo H1. Describir los principales rasgos de las actividades formativas realizadas por el profesorado en los últimos años con relación a la evaluación del aprendizaje.	575
5.2 Resultados Objetivo H2. Identificar las principales necesidades formativas manifestadas por el profesorado con relación al desarrollo de su competencia evaluativa.	578
6. El capítulo a grandes rasgos.....	582

CAPÍTULO 9. COMPETENCIA EVALUATIVA DEL PROFESORADO EN LA PRÁCTICA: CONCLUSIONES DEL MARCO EMPÍRICO 597

1. Objetivo E. Realizar una autoevaluación de la competencia evaluativa del profesorado dedicado a la formación básica de maestras/os de Educación Infantil y/o Primaria en el contexto institucional de la UD y BAM	598
1.1 Factores e indicadores de la evaluación del aprendizaje mejor y peor valorados por la muestra docente y su reflejo en la investigación y la literatura (Objetivo E1).	598
1.2 Relación entre variables socio-demográficas, académicas y profesionales y la competencia evaluativa autopercebida: conclusiones y reflejo de las mismas en la investigación y la literatura (Objetivo E2).....	609
1.3 Obstáculos y apoyos en el ejercicio de la evaluación según la muestra docente y su reflejo en la investigación y la literatura (Objetivo E3).	612
2. Objetivo F. Realizar un diagnóstico de la competencia evaluativa del profesorado dedicado a la formación básica de maestras/os de Educación Infantil y/o Primaria en la UD y BAM desde la perspectiva de las/os estudiantes.....	614
2.1 Factores e indicadores de la evaluación del aprendizaje mejor y peor valorados en la muestra discente y su reflejo en la investigación y la literatura (Objetivo F1).	614
2.2 Relación entre variables socio-demográficas, académicas y profesionales y la competencia evaluativa docente percibida por el alumnado: conclusiones y reflejo de las mismas en la investigación y la literatura (Objetivo F2).	627
2.3 Obstáculos y apoyos en el ejercicio de la evaluación según la muestra discente y su reflejo en la investigación y la literatura (Objetivo F3).	629
3. Objetivo G. Contrastar el punto de vista que manifiestan docentes y discentes con respecto a la competencia evaluativa del profesorado vinculado a la formación básica de maestras/os en la UD y BAM, indagando tanto en la perspectiva teórica (importancia), como en la práctica (realización)	631
4. Objetivo H. Describir las experiencias y demandas que plantea el profesorado con respecto a la formación permanente vinculada a la evaluación del aprendizaje	637
MARCO CONCLUSIVO.....	645
CAPÍTULO 10. CONCLUSIONES DE LA INVESTIGACIÓN	647
1. Conclusiones: claves centrales de la discusión y sugerencias de mejora.....	649
2. Contribuciones de la investigación	659
3. Dificultades, limitaciones y sugerencias de mejora.....	661
4. Líneas de investigación	667
REFERENCIAS BIBLIOGRÁFICAS.....	673

REFERENCIAS BIBLIOGRÁFICAS	675
ESCALAS DE EVALUACIÓN DE LA CALIDAD DOCENTE	786
ANEXOS ELECTRÓNICOS	en CD adjunto

ÍNDICE DE TABLAS

Tabla 1	<i>La investigación de la competencia evaluativa en la formación básica de maestras/os. Evidencias indirectas para su cuantificación</i>	38
Tabla 2	<i>Principales cambios pedagógicos y estructurales vinculados al proceso de Bolonia</i>	59
Tabla 3	<i>Factores sociales y académico-institucionales que dificultan el ejercicio docente</i>	72
Tabla 4	<i>Modelo de competencias docentes del profesorado de la Universidad de Deusto (Villa & García, 2008; cit. en Villa, 2008, pp. 201-203)</i>	84
Tabla 5	<i>Correspondencia de los estándares de calidad y excelencia docente de la UD y los perfiles profesionales analizados</i>	87
Tabla 6	<i>Principales rasgos de las cuatro generaciones de la evaluación (Guba & Lincoln, 1989)</i>	110
Tabla 7	<i>Principales conceptualizaciones dentro del paradigma de la indagación: reinterpretación de la propuesta de Falchikov (2005)</i>	117
Tabla 8	<i>Comparación de los enfoques evaluativos tradicional y alternativo. Revisión ampliada de la propuesta de Wolf y Reader (1996)</i>	121
Tabla 9	<i>Representación de las claves de la definición evaluativa adoptada en trabajos previos de síntesis y otras conceptualizaciones</i>	129
Tabla 10	<i>Secuencia interna de pasos en la planificación evaluativa</i>	132
Tabla 11	<i>Secuencia interna de pasos en la implementación o puesta en práctica evaluativa</i>	133
Tabla 12	<i>Secuencia interna de pasos en la planificación evaluativa</i>	135
Tabla 13	<i>Propósitos específicos vinculados a la función social de la evaluación del aprendizaje</i>	138
Tabla 14	<i>Propósitos específicos vinculados a la función pedagógica de la evaluación</i>	139
Tabla 15	<i>Clasificación revisada de las principales tipologías aplicables a la evaluación del aprendizaje</i>	143
Tabla 16	<i>Antecedentes en el análisis de la praxis evaluativa</i>	152
Tabla 17	<i>Síntomas de disfuncionalidad en la evaluación del aprendizaje universitario</i>	155
Tabla 18	<i>Descriptorios generales de los sistemas tradicionales de evaluación del aprendizaje en la universidad</i>	171
Tabla 19	<i>Efectos de los principales rasgos de disfuncionalidad evaluativa en el alumnado universitario: aprendizajes implícitos sobre la evaluación</i>	176
Tabla 20	<i>Efectos de los principales rasgos de disfuncionalidad evaluativa en el profesorado universitario: aprendizajes implícitos sobre la evaluación que generan desprofesionalización</i>	179
Tabla 21	<i>Resumen de los rasgos básicos de las concepciones evaluativas profundas</i>	192
Tabla 22	<i>Descripción de los rasgos básicos de las concepciones evaluativas profundas</i>	193
Tabla 23	<i>Resistencias al cambio de las concepciones evaluativas profundas: descripción de causas</i>	194
Tabla 24	<i>Rasgos de la formación inicial y permanente del profesorado universitario para la mejora de los sistemas de evaluación del aprendizaje</i>	202
Tabla 25	<i>Rasgos característicos de la evaluación del aprendizaje en el contexto específico de la formación básica de maestras/os</i>	213
Tabla 26	<i>Principales rasgos de la evaluación para el aprendizaje</i>	228

Tabla 27 <i>Implicaciones del EEES en la evaluación del aprendizaje discente: resumen de aportaciones</i>	230
Tabla 28 <i>Exigencias evaluativas del EEES: conclusiones para la definición de una teoría de la evaluación del aprendizaje contextualizada</i>	233
Tabla 29 <i>Referentes normativos para el ejercicio de la evaluación: conclusiones para la definición de una teoría de la evaluación del aprendizaje contextualizada</i>	234
Tabla 30 <i>Directrices para el desarrollo de sistemas de garantía interna de la calidad del programa AUDIT. Descriptores referidos a la evaluación del aprendizaje (ANECA, AQU, & ACSUC, 2007b)</i>	237
Tabla 31 <i>Implicaciones evaluativas en la formación básica de maestras/os: conclusiones para la definición de una teoría de la evaluación del aprendizaje contextualizada</i>	239
Tabla 32 <i>La competencia evaluativa en los decretos reguladores para la formación básica de maestras/os de Educación Infantil y Primaria</i>	242
Tabla 33 <i>Antecedentes en la sistematización de un modelo o teoría de la evaluación del aprendizaje</i>	245
Tabla 34 <i>Antecedentes en la sistematización de rasgos evaluativos deseables: conclusiones para la definición de una teoría de la evaluación del aprendizaje fundamentada</i>	246
Tabla 35 <i>Elementos clave para la definición de una teoría de la evaluación: análisis de los antecedentes teóricos</i>	251
Tabla 36 <i>Rasgos teóricos de la buena evaluación: conclusiones comunicadas en otras propuestas</i>	253
Tabla 37 <i>Elementos clave para la definición de una teoría de la evaluación: análisis de las propuestas estructuradas</i>	257
Tabla 38 <i>Propuesta de estándares para la evaluación de los aprendizajes discentes (JCSEE, 2007, pp. 4-5)</i>	260
Tabla 39 <i>Modelo ITC: subprocesos y criterios</i>	265
Tabla 40 <i>Características básicas de las competencias como objeto evaluativo específico</i>	273
Tabla 41 <i>Implicaciones de la ampliación del objeto de evaluación en el EEES: resumen comparativo (Bilbao, 2011b, pp. 115-116)</i>	274
Tabla 42 <i>Características principales de la evaluación contextualizada</i>	277
Tabla 43 <i>Clasificación de las principales modalidades de feedback</i>	297
Tabla 44 <i>Dimensiones de la competencia evaluativa del profesorado universitario en los grados dedicados a la formación básica de maestras/os de Educación Infantil y Primaria</i>	327
Tabla 45 <i>Antecedentes en la sistematización de un modelo o teoría de la evaluación del aprendizaje</i>	343
Tabla 46 <i>Antecedentes en la sistematización de la competencia evaluativa del profesorado en la formación básica de maestras/os (FB), en la docencia universitaria (DU) o en la enseñanza general</i>	343
Tabla 47 <i>Referentes para la configuración del modelo de competencia evaluativa docente: escalas y cuestionarios aplicados al análisis de las prácticas evaluativas en la universidad</i>	349
Tabla 48 <i>Representación de los principales tópicos evaluativos en las dimensiones integrantes del modelo</i>	358
Tabla 49 <i>Principales componentes integrados en la planificación estratégica de la evaluación para el aprendizaje (dimensión 01)</i>	366
Tabla 50 <i>Principales componentes integrados en la explicitación y transparencia de la evaluación (dimensión 02)</i>	370
Tabla 51 <i>Principales componentes integrados en la diversificación de la evaluación (dimensión 03)</i>	373

Tabla 52 Principales componentes integrados en la implementación ajustada de la evaluación (dimensión 04)	376
Tabla 53 Principales componentes integrados en la interpretación y análisis de evidencias (dimensión 05)	380
Tabla 54 Principales componentes integrados en la regulación del aprendizaje y la enseñanza (dimensión 06)	385
Tabla 55 Principales componentes integrados en la interpretación y análisis de evidencias (dimensión 07)	388
Tabla 56 Principales componentes integrados en la revisión del dispositivo evaluativo (dimensión 08)	392
Tabla 57 Objetivos de la investigación abordados en el marco empírico	409
Tabla 58 Ventajas y limitaciones de los rasgos del diseño de investigación ideado.....	413
Tabla 59 Síntesis de la representatividad de la población en la muestra discente, según datos facilitados por la Secretaría Académica de la Facultad de Psicología y Educación de la UD y de BAM.....	419
Tabla 60 Población (N) y tamaño de la muestra (n) según orientaciones de Carmona (2002)	422
Tabla 61 Clasificación de las variables contempladas en el marco de la investigación	423
Tabla 62 Principales amenazas en los procesos de evaluación de la calidad docente con incidencia potencial en la evaluación o autoevaluación de la competencia evaluativa.....	431
Tabla 63 Estructuración del cuestionario de autoevaluación de la competencia evaluativa docente	434
Tabla 64 Estructuración del cuestionario de evaluación de la competencia evaluativa docente para estudiantes.....	435
Tabla 65 Proceso de creación del cuestionario en la versión docente	437
Tabla 66 Extracción de componentes principales y varianza total explicada en la subescala de importancia en docentes.....	441
Tabla 67 Matriz de componentes rotados en la subescala de importancia en docentes	442
Tabla 68 Orientaciones para interpretar el tamaño de la relación según Bisquerra (1987) y Morales (2008a).....	443
Tabla 69 Componentes del Factor 1 en la versión docente del cuestionario (Planificación estratégica de la evaluación) y sus cargas factoriales	445
Tabla 70 Componentes del Factor 2 en la versión docente del cuestionario (Regulación del aprendizaje y la enseñanza por medio de la evaluación) y sus cargas factoriales	446
Tabla 71 Componentes del Factor 3 en la versión docente del cuestionario (Metaevaluación) y sus cargas factoriales.....	446
Tabla 72 Componentes del Factor 4 en la versión docente del cuestionario (Ajustes de la evaluación y participación discente) y sus cargas factoriales.....	447
Tabla 73 Componentes del Factor 5 en la versión docente del cuestionario (Adecuación y precisión) y sus cargas factoriales	448
Tabla 74 Componentes del Factor 6 en la versión docente del cuestionario (Puesta en práctica de la evaluación) y sus cargas factoriales.....	449
Tabla 75 Extracción de componentes principales y varianza total explicada en la subescala de importancia en estudiantes	449
Tabla 76 Matriz de componentes rotados en la subescala de importancia en estudiantes	452
Tabla 77 Componentes del Factor 1 en la versión discente del cuestionario (Participación del alumnado en la evaluación) y sus cargas factoriales	456
Tabla 78 Componentes del Factor 2 en la versión discente del cuestionario (Planificación del sistema evaluativo) y sus cargas factoriales.....	457

Tabla 79 Componentes del Factor 3 en la versión discente del cuestionario (Ajustes en el sistema evaluativo) y sus cargas factoriales	457
Tabla 80 Componentes del Factor 5 en la versión discente del cuestionario (Regulación del aprendizaje por medio de la evaluación) y sus cargas factoriales	458
Tabla 81 Componentes del Factor 5 en la versión discente del cuestionario (Adecuación y precisión) y sus cargas factoriales.....	459
Tabla 82 Componentes del Factor 6 en la versión discente del cuestionario (Control de las condiciones en las que se desarrolla la evaluación) y sus cargas factoriales.....	459
Tabla 83 Correlación y covariación de los componentes del factor sexto en la subescala de realización del cuestionario docente.....	464
Tabla 84 Correlación y covariación de los componentes del factor sexto en la subescala de importancia del cuestionario discente.....	466
Tabla 85 Valoración de las posibilidades y limitaciones asociadas a las herramientas de recogida de información.....	467
Tabla 86 Síntesis del proceso de recogida de información desarrollado	471
Tabla 87 Supuestos de aplicación de las técnicas estadísticas paramétricas	474
Tabla 88 Análisis tentativo para la valoración del supuesto de normalidad en la muestra	478
Tabla 89 Pruebas de inferencia estadística no paramétrica empleadas.....	479
Tabla 90 Resumen de los principales rasgos del marco empírico de la investigación	481
Tabla 91 Distribución de la muestra docente por titulación de acceso (formación previa)	485
Tabla 92 Valoración de la muestra docente en los seis factores asociados a la competencia evaluativa	489
Tabla 93 Conclusiones estadísticamente significativas ($p < .05$) en la valoración de la competencia evaluativa efectuada en la muestra docente por factor y subescala (Sub.).....	494
Tabla 94 Autovaloración de la competencia evaluativa en la muestra docente en función del SEXO	497
Tabla 95 Diferencias significativas en las valoraciones de la muestra docente en función del SEXO: estadísticos de grupo	497
Tabla 96 Autovaloración de la competencia evaluativa en la muestra de profesoras/es en función del NIVEL DE EXPERIENCIA DOCENTE.....	498
Tabla 97 Diferencias significativas en las valoraciones de la muestra de profesoras/es en función del NIVEL DE EXPERIENCIA DOCENTE: estadísticos de grupo y contraste post hoc (Scheffé)	499
Tabla 98 Autovaloración de la competencia evaluativa en la muestra docente en función de la FORMACIÓN PREVIA DEL PROFESORADO.....	500
Tabla 99 Formación previa del profesorado y autoevaluación de su competencia evaluativa	502
Tabla 100 Autovaloración de la competencia evaluativa en la muestra docente en función de la SIMULTANEIDAD o EXCLUSIVIDAD en el ejercicio de la enseñanza.....	503
Tabla 101 Diferencias significativas en las valoraciones de la muestra docente en función de la SIMULTANEIDAD (SI) o EXCLUSIVIDAD (NO) en el ejercicio de la enseñanza: estadísticos de grupo	504
Tabla 102 Autovaloración de la competencia evaluativa en la muestra docente en función del CENTRO	505
Tabla 103 Diferencias significativas en las valoraciones de la muestra docente en función del CENTRO: estadísticos de grupo.....	505
Tabla 104 Autovaloración de la competencia evaluativa en la muestra docente en función del CURSO	506

Tabla 105 <i>Diferencias significativas en las valoraciones de la muestra docente en función del CURSO: estadísticos de grupo y contraste post hoc (Scheffé y DMS)</i>	507
Tabla 106 <i>Promedios de valoración en la muestra docente en función de la COMPETENCIA EVALUATIVA AUTOPERCIBIDA (poco competente y competente/muy competente)</i>	508
Tabla 107 <i>Diferencias significativas en las valoraciones de la muestra docente en función de la COMPETENCIA EVALUATIVA AUTOPERCIBIDA</i>	509
Tabla 108 <i>Autovaloración de la competencia evaluativa en la muestra docente en función de la FORMACIÓN PERMANENTE EN EVALUACIÓN</i>	510
Tabla 109 <i>Diferencias significativas en las valoraciones de la muestra docente en función de la FORMACIÓN PERMANENTE EN EVALUACIÓN: estadísticos de grupo</i>	511
Tabla 110 <i>Autovaloración de la competencia evaluativa en la muestra docente en función de la AUTOPERCEPCIÓN DE NECESIDAD FORMATIVA</i>	511
Tabla 111 <i>Diferencias significativas en las valoraciones de la muestra docente en función de la NECESIDAD FORMATIVA AUTOPERCIBIDA: estadísticos de grupo</i>	512
Tabla 112 <i>Apoyos para el ejercicio de la competencia evaluativa según las respuestas abiertas del profesorado</i>	515
Tabla 113 <i>Obstáculos para el ejercicio de la competencia evaluativa según las respuestas abiertas del profesorado</i>	516
Tabla 114 <i>Valoración de la muestra discente en los seis factores asociados a la competencia evaluativa docente</i>	520
Tabla 115 <i>Conclusiones estadísticamente significativas ($p < .05$) en la valoración de la competencia evaluativa docente en la muestra discente por factor y subescala (Sub.)</i>	525
Tabla 116 <i>Valoración de la competencia evaluativa docente en la muestra discente en función del SEXO</i>	529
Tabla 117 <i>Diferencias significativas en las valoraciones de la muestra discente en función del SEXO: estadísticos de grupo</i>	530
Tabla 118 <i>Correlación de la edad del alumnado y la valoración de la competencia evaluativa docente</i>	531
Tabla 119 <i>Valoración de la competencia evaluativa docente en la muestra discente en función de la EDAD</i>	532
Tabla 120 <i>Diferencias significativas en la valoración de la competencia evaluativa docente en función de la EDAD del alumnado: estadísticos de grupo y contraste post hoc (Bonferroni y Scheffé)</i>	533
Tabla 121 <i>Valoración de la competencia evaluativa docente en la muestra discente en función de la DEDICACIÓN A LA FORMACIÓN UNIVERSITARIA</i>	534
Tabla 122 <i>Diferencias significativas en las valoraciones de la muestra docente en función de la DEDICACIÓN A LA FORMACIÓN: estadísticos de grupo y contraste post hoc (Bonferroni, DMS y Scheffé)</i>	535
Tabla 123 <i>Valoración de la competencia evaluativa docente en la muestra discente en función del GRADO</i>	536
Tabla 124 <i>Diferencias significativas en las valoraciones de la muestra discente en función del GRADO: estadísticos de grupo</i>	536
Tabla 125 <i>Valoración de la competencia evaluativa docente en la muestra discente en función del CURSO</i>	537
Tabla 126 <i>Diferencias significativas en las valoraciones de la muestra docente en función del CURSO: estadísticos de grupo y contraste post hoc (Bonferroni, DMS y Scheffé)</i>	538
Tabla 127 <i>Valoración de la competencia evaluativa docente en la muestra discente en función del CENTRO</i>	539

Tabla 128 <i>Diferencias significativas en las valoraciones de la muestra discente en función del CENTRO: estadísticos de grupo</i>	540
Tabla 129 <i>Valoración de la competencia evaluativa docente en la muestra discente en función de la UNIDAD de REFERENCIA (módulo o asignatura)</i>	541
Tabla 130 <i>Diferencias significativas en las valoraciones de la muestra discente en función de la UNIDAD de DOCENCIA valorada: estadísticos de grupo</i>	542
Tabla 131 <i>Valoración de la competencia evaluativa docente en la muestra discente en función de la MATERIA</i>	543
Tabla 132 <i>Valoración de la competencia evaluativa docente en la muestra discente en función del MÓDULO</i>	544
Tabla 133 <i>Diferencias significativas en la valoración de la competencia evaluativa docente en función del MÓDULO: estadísticos de grupo y contraste post hoc (T2 de Tamhane, T3 de Dunnet, C de Dunnet, Games-Howel, Bonferroni y Scheffé)</i>	545
Tabla 134 <i>Valoración de la competencia evaluativa docente en la muestra discente en función de la EXPERIENCIA PREVIA</i>	548
Tabla 135 <i>Diferencias significativas en las valoraciones de la muestra discente en función de la EXPERIENCIA PREVIA: estadísticos de grupo</i>	549
Tabla 136 <i>Valoración de la competencia evaluativa docente en la muestra discente en función de los RESULTADOS PARCIALES</i>	549
Tabla 137 <i>Diferencias significativas en las valoraciones de la muestra discente en función de los RESULTADOS PARCIALES: estadísticos de grupo</i>	550
Tabla 138 <i>Valoración de la competencia evaluativa docente en la muestra discente en función de las EXPECTATIVAS ACADÉMICAS</i>	551
Tabla 139 <i>Diferencias significativas en la valoración de la competencia evaluativa docente en función de las EXPECTATIVAS ACADÉMICAS del alumnado: estadísticos de grupo y contraste post hoc (T2 de Tamhane, T3 de Dunnet, C de Dunnet, Games-Howel, Bonferroni y Scheffé)</i> ..	552
Tabla 140 <i>Correlación de la edad del alumnado y la valoración de la competencia evaluativa docente</i>	553
Tabla 141 <i>Valoración de la competencia evaluativa docente en la muestra discente en función de la COMPETENCIA DOCENTE ATRIBUIDA</i>	554
Tabla 142 <i>Diferencias significativas en la valoración de la competencia evaluativa docente en función de la COMPETENCIA DOCENTE ATRIBUIDA por el alumnado: estadísticos de grupo y contraste post hoc (T2 de Tamhane, T3 de Dunnet, C de Dunnet, Games-Howel, Bonferroni y Scheffé)</i>	555
Tabla 143 <i>Valoración de la competencia evaluativa docente en la muestra discente en función de la MODALIDAD DE RESPUESTA</i>	556
Tabla 144 <i>Diferencias significativas en las valoraciones de la muestra discente en función de la MODALIDAD DE RESPUESTA: estadísticos de grupo</i>	557
Tabla 145 <i>Fortalezas en la evaluación del aprendizaje según el alumnado de la muestra</i>	560
Tabla 146 <i>Áreas de mejora en la evaluación del aprendizaje según las respuestas abiertas del alumnado</i>	562
Tabla 147 <i>Contraste de promedios de importancia de estudiantes y profesorado en las dimensiones comparables de la competencia evaluativa</i>	569
Tabla 148 <i>Diferencias significativas en el contraste de los promedios de importancia comparables en función de la FUENTE: estadísticos de grupo</i>	569
Tabla 149 <i>Contraste de promedios de importancia de estudiantes y profesorado en los componentes de la Regulación del aprendizaje por medio de la evaluación (D5 – II)</i>	570
Tabla 150 <i>Diferencias significativas en los componentes de la Regulación del aprendizaje por medio de la evaluación (D5 – II) en función de la FUENTE: estadísticos de grupo</i>	570

Tabla 151	<i>Contraste de promedios de realización percibida en estudiantes y profesorado para las dimensiones comparables de la competencia evaluativa.....</i>	571
Tabla 152	<i>Diferencias significativas en el contraste de los promedios de realización de las dimensiones comparables en función de la FUENTE: estadísticos de grupo.....</i>	572
Tabla 153	<i>Contraste de promedios de realización percibida en estudiantes y profesorado para los indicadores comparables de la competencia evaluativa</i>	572
Tabla 154	<i>Diferencias significativas en el contraste de los promedios de realización de los indicadores en función de la FUENTE: estadísticos de grupo</i>	573
Tabla 155	<i>Principales modalidades empleadas en la formación permanente evaluativa del profesorado de la muestra</i>	576
Tabla 156	<i>Principales temas abordados en la formación permanente evaluativa del profesorado de la muestra</i>	577
Tabla 157	<i>Principales modalidades de formación permanente sugeridas por el profesorado para la capacitación evaluativa.....</i>	578
Tabla 158	<i>Principales temas de formación permanente sugeridos por el profesorado para la capacitación evaluativa.....</i>	579
Tabla 159	<i>Competencia evaluativa autopercebida y variables académicas, socio-demográficas o profesionales: síntesis de resultados (objetivo E2)</i>	584
Tabla 160	<i>Competencia evaluativa percibida por el alumnado en función de variables académicas, socio-demográficas o profesionales: síntesis de resultados (objetivo F2)</i>	588
Tabla 161	<i>Competencia evaluativa percibida por el alumnado en función de variables académicas, socio-demográficas o profesionales: síntesis de resultados (objetivo G).....</i>	593
Tabla 162	<i>Repuesta articulada a las preguntas y objetivos de la investigación</i>	648

ÍNDICE DE FIGURAS

<i>Figura 1.</i> Fases recorridas en el proceso de investigación.	47
<i>Figura 2.</i> Teoría de la competencia evaluativa docente: Estructura interna del marco teórico.....	49
<i>Figura 3.</i> La competencia evaluativa docente en la práctica: estructura del marco empírico de la investigación y conclusiones.....	51
<i>Figura 4.</i> El proceso de Bolonia: cronograma básico. Elaboración propia a partir de Comas (2013), European Higher Education Area (EHEA, 2010) y Estrategia Europa 2020 (Comisión Europea, 2010).	57
<i>Figura 5.</i> La competencia evaluativa del profesorado universitario: claves para su delimitación.	89
<i>Figura 6.</i> Razones para profundizar sobre la competencia evaluativa del profesorado.	92
<i>Figura 7.</i> La competencia evaluativa del profesorado universitario de las titulaciones dedicadas a la formación de maestros como elemento estratégico para la mejora.	105
<i>Figura 8.</i> Las generaciones evolutivas del concepto evaluación y su relación con los paradigmas de evaluación y los enfoques evaluativos.....	113
<i>Figura 9.</i> Paradigmas de la evaluación, conceptualizaciones asociadas y principales enfoques.....	120
<i>Figura 10.</i> El proceso de evaluación: propuesta flexible de fases y tareas.....	136
<i>Figura 11.</i> Evaluación del aprendizaje universitario: análisis del estado de la cuestión.	150
<i>Figura 12.</i> Causas profundas y estructurales de disfuncionalidad evaluativa.....	182
<i>Figura 13.</i> Propuestas para la mejora de la evaluación del aprendizaje en la universidad.	197
<i>Figura 14.</i> Ciclo de reproducción de las prácticas evaluativas disfuncionales por medio de la formación básica de maestras/os.	214
<i>Figura 15.</i> Referencias para la construcción de una teoría de la evaluación (modelo ITC).....	223
<i>Figura 16.</i> Delimitación conceptual de la evaluación para el aprendizaje en el marco de la investigación en curso.....	225
<i>Figura 17.</i> Ampliación de la evaluación del aprendizaje en el contexto formativo del EEES.....	232
<i>Figura 18.</i> Modelo ITC (Integrado Teórico Complejo) de evaluación.	263
<i>Figura 19.</i> Modalidades de evaluación participativa contempladas en el modelo ITC.....	280
<i>Figura 20.</i> Competencia evaluativa del profesorado: conclusión del marco teórico.	325
<i>Figura 21.</i> Proceso de definición del modelo de competencia evaluativa docente.	336
<i>Figura 22.</i> Definición de la rúbrica asociada al modelo de competencia evaluativa: descripción del proceso.	340
<i>Figura 23.</i> Referentes para la construcción de un modelo de sistematización de la competencia evaluativa docente.	341
<i>Figura 24.</i> Representación de las 15 categorías temáticas identificadas en el banco de tópicos evaluativos.	355
<i>Figura 25.</i> Resumen del proceso de muestreo desarrollado según Fox (1981).	416
<i>Figura 26.</i> La incidencia de la accidentalidad en la muestra docente y discente.....	418
<i>Figura 27.</i> Configuración de la muestra discente por estratos: centro (BAM y UD), titulación (EI- Educación Infantil y EP- Educación Primaria) y curso (1º, 2º y 3º).	420
<i>Figura 28.</i> Configuración de la muestra docente por estratos: centro (BAM y UD), titulación (EI- Educación Infantil y EP- Educación Primaria) y curso (1º, 2º y 3º).	421

<i>Figura 29.</i> La creación de las herramientas de recogida de información en el marco general de la investigación.	436
<i>Figura 30.</i> Estructura interna de la respuesta docente y relación con el modelo teórico de referencia.	444
<i>Figura 31.</i> Estructura interna de la respuesta discente y relación con el modelo teórico de referencia.	454
<i>Figura 32.</i> Índices de consistencia interna en los factores y subescalas del cuestionario docente. ...	462
<i>Figura 33.</i> Índices de consistencia interna en los factores y subescalas del cuestionario discente. .	465
<i>Figura 34.</i> Distribución de la muestra docente por años de experiencia en la enseñanza universitaria.	484
<i>Figura 35.</i> Formación de postgrado en la muestra docente.	485
<i>Figura 36.</i> Distribución de la muestra de estudiantes por grupos de edad.	486
<i>Figura 37.</i> Dedicación de la muestra de estudiantes a su propia formación universitaria.	487
<i>Figura 38.</i> Representación AIR de los factores asociados a la competencia evaluativa docente en la muestra de profesoras/es.	490
<i>Figura 39.</i> Promedios de importancia y realización declarados por la muestra docente en los 19 indicadores de la competencia evaluativa docente.	491
<i>Figura 40.</i> Resumen del tono general de las respuestas docentes en la pregunta referida a los apoyos y obstáculos en el ejercicio de la competencia evaluativa.	514
<i>Figura 41.</i> Representación AIR de los factores asociados a la competencia evaluativa docente en la muestra de estudiantes.	521
<i>Figura 42.</i> Promedios de importancia y realización declarados por la muestra discente en los 20 indicadores de competencia evaluativa.	522
<i>Figura 43.</i> Resumen del tono general de las respuestas del alumnado en la pregunta referida a las fortalezas y las áreas de mejora en la evaluación de su aprendizaje.	559
<i>Figura 44.</i> Representación de las tres principales fortalezas evaluativas señaladas por la muestra discente con los datos recabados por medio del cuestionario.	561
<i>Figura 45.</i> Perspectiva docente y discente en los componentes comparables en la valoración de la competencia evaluativa.	567
<i>Figura 46.</i> Satisfacción de la muestra docente con la formación permanente evaluativa desarrollada en los dos últimos años.	576

PLANTEAMIENTO GENERAL DE LA INVESTIGACIÓN

CAPÍTULO INTRODUCTORIO

El objetivo de este apartado inicial es ofrecer una visión general del trabajo de investigación presentado para quien se aproxima a él por primera vez. Para anticipar el desarrollo previsto en los próximos capítulos, realiza un recorrido sintético por cuatro apartados principales:

- La elección del tema central del estudio y su interés, relevancia y pertinencia.
- Las preguntas centrales que guían la investigación y los objetivos formulados para su resolución.
- El proceso de investigación desarrollado para la consecución de dichos propósitos.
- La estructuración de la memoria de investigación elaborada para dejar constancia de dicho recorrido y de las principales conclusiones obtenidas.

1. Justificación: relevancia, interés y pertinencia de la investigación

La presente investigación centra su atención en la competencia evaluativa del profesorado universitario dedicado a la formación básica de maestras/os de Educación Infantil y Primaria. Enmarcada en el área de la Didáctica universitaria, su desarrollo responde a la última de las cinco tendencias temáticas que sintetiza Gros (2006) para la investigación actual sobre docencia universitaria. De acuerdo con la misma, plantea un estudio sobre los sistemas de evaluación del aprendizaje que diseña y pone en práctica el profesorado, concebidos como un recurso para la estimulación consciente del aprendizaje discente y su acreditación.

La aproximación a la competencia evaluativa de este sector específico del profesorado universitario está motivada por la confluencia de cuatro fuentes de influencia que otorgan al tema interés y pertinencia:

- La primera de ellas, es la certeza ***del importante papel que desempeña la evaluación con respecto al aprendizaje universitario***. La investigación específica lleva décadas señalando el influjo que ejerce la evaluación sobre las estrategias de aprendizaje discente y, en consecuencia, sobre la calidad del aprendizaje mismo (Biggs, 2005; Entwistle, 1984/2005; Gibbs & Dunbar-Goddet, 2007; Hernández, 1996; Marton & Säljö, 1984/2005; Meyer & Watson, 1991). Aunque tradicionalmente haya sido concebida y empleada como instrumento de acreditación, el poder regulador que ejerce la evaluación (Barberá, 1999) hace de ella un recurso didáctico privilegiado para promover deliberadamente el logro de los objetivos de aprendizaje establecidos (Carless, 2007; Carless, Joughin, & Mok, 2006; J. Chappuis, 2009; S. Chappuis & Stiggins, 2002). En el alumnado genera un estímulo y un poder movilizador que favorece la orientación estratégica del trabajo y el esfuerzo en línea con las exigencias evaluativas (Arribas, 2012; Gibbs & Simpson, 2009; Monereo, 2003; Román & Murillo, 2009); en el profesorado, ofrece una referencia clara desde la que articular y dirigir el proceso formativo de principio a fin (Parcerisa, 2010; Quesada, Rodríguez, & Ibarra, 2013; Shepard, 2000). En consecuencia, la evaluación se convierte en el elemento clave que condiciona la consolidación práctica de los procesos de innovación y mejora que deseen emprenderse sobre la enseñanza, el currículo y/o el aprendizaje (Bonsón & Benito, 2005; Dochy, Segers, & Diederick, 2002).
- A pesar de su manifiesta y reconocida importancia, ***la evaluación del aprendizaje en el contexto universitario español está repleta de síntomas de disfuncionalidad y deficiencia***, y ello representa una segunda motivación para profundizar en la competencia evaluativa docente. Estudios efectuados a lo largo de las dos últimas décadas revelan el predominio de prácticas alejadas del planteamiento didáctico-pedagógico de la evaluación (M. Fernández, 1989; Nieto, 2000; Padilla, Gil, Rodríguez, & Clarés, 2010; G. Rodríguez, Ibarra, Gallego, Gómez, & Quesada, 2012; Tejedor, 1999; Trillo & Porto, 1999b, dentro de una extensa lista de trabajos publicados). Las manifestaciones de disfuncionalidad que se señalan, no sólo se producen en la educación superior, sino que tienen lugar desde los niveles educativos iniciales (M. A. Calatayud, 2004; D. Salinas, 2002; Santos, 2003). De lo anterior se deduce la necesidad de investigar para mejorar las prácticas evaluativas que lleva a cabo el profesorado y, con ello, la enseñanza y el aprendizaje.
- La emergencia del Espacio Europeo de Educación Superior (EEES) representa un tercer impulso en la elección del tema. Admitida la relevancia del fenómeno

evaluativo y las dificultades que entraña su puesta en práctica, la emergencia de un nuevo paradigma formativo representa una oportunidad para avanzar en la redefinición y exploración de nuevas formas de pensar y hacer evaluación (A. Calatayud, 2008; Rué, 2008; Villardón, 2006; Yániz, 2008). Las directrices que emanan del proceso de convergencia dan lugar a un nuevo escenario universitario y a la reinterpretación de los roles discente y docente dentro del mismo. En respuesta a esta nueva situación, se constata **la necesidad de nuevos modelos evaluativos, coherentes con el planteamiento pedagógico de base**. Ello se traduce en toda una serie de demandas que afectan a las finalidades, los objetos, los momentos, el realismo, la complejidad o los agentes de la evaluación (M. J. García, 2010; Mateo, 2006b, 2008; Monescillo & Toscano, 2007; B. Salinas & Cotillas, 2007; J. Sánchez, 2011), indispensables para que se resuelva con éxito el proyecto de la formación universitaria europea (Buscà, Martínez, Pintor, & Peire, 2010; Leal, 2004; Olmos & Rodríguez, 2008; Ureña, Valles, & Ruiz, 2009).

- La cuarta fuente de inspiración e influencia se detiene en la **particular relevancia que adopta la evaluación del aprendizaje universitario en aquellas titulaciones dedicadas a la formación básica de maestras/os**. La forma de plantear y llevar a cabo la evaluación del aprendizaje, depende de las concepciones profundas que se sostengan sobre la misma (Buendía, Carmona, González, & López, 1999; De la Cruz, Pozo, Huarte, & Scheuer, 2006; Korthagen, 2010). En ellas es bien conocido el efecto de las experiencias acumuladas desde la perspectiva discente (Biggs, 2005; Feixas, 2002, 2008; Monereo & Castelló, 2009a), que generan una potente tendencia de reproducción inconsciente al comenzar a ejercer la enseñanza (B. Álvarez, González, & García, 2007; V. M. López, 2004). Este círculo de influencias hace que la evaluación del aprendizaje adopte una importancia capital en estas titulaciones (C. Brew, Relye, & Walta, 2009; V. M. López, 2008; V. M. López, Manrique, & Vallés, 2011; V. M. López & Palacios, 2012; Palacios & López, 2013). Prácticas evaluativas orientadas a la promoción y el fortalecimiento del aprendizaje, y no sólo su medición, demostrarían a las/os futuras/os evaluadoras/es que su empleo pedagógico es posible; por contra, prácticas evaluativas meramente acreditativas y/o disfuncionales reafirmarían la visión negativa que rodea este proceso (L. F. Martínez, Castejón, & Santos, 2012; Porto & Rodríguez, 2009; Ricoy & Fernández-Rodríguez, 2013), alimentando su probabilidad de perpetuación desde los niveles básicos del sistema educativo. Desde esta perspectiva, advertimos en el desarrollo de la competencia evaluativa del profesorado un reto general de mejora

asociado a la formación universitaria, y una línea de intervención prioritaria en las titulaciones dedicadas a la preparación de las/os futuras/os maestras/os.

La consideración conjunta y consecuente de las cuatro influencias reseñadas, lleva a centrar la investigación en el ámbito de las titulaciones vinculadas a la formación básica de maestras y maestros de Educación Infantil y Primaria en el contexto español. Tal delimitación supone apartar del foco de atención los estudios de máster que actualmente dan acceso a la enseñanza secundaria en todo el estado. Aunque el análisis de la competencia evaluativa resultaría igualmente pertinente en dicho caso, su adscripción a los estudios de postgrado introduce una serie de especificidades difíciles de gestionar desde la perspectiva de la investigación. Orientada a la especialización, su estructuración es mucho más flexible y diversa que en los grados de formación básica (extensión desigual que oscila entre 60 y 120 ECTS frente a los 240 ECTS que conforman el grado, combinaciones dispares de modalidades formativas presencial y no presencial, etc.), por lo que su inclusión supondría asumir una importante cantidad de factores particulares, difíciles de predecir y controlar.

El interés general que suscita la evaluación del aprendizaje se hace manifiesto en el creciente número de estudios y publicaciones que sitúan en ella su foco de atención (I. Álvarez, 2008b; Ibarra & Rodríguez, 2010). Dos búsquedas efectuadas a modo groso ofrecen una evidencia cuantitativa de ello:

- La introducción de los términos *assessment* y *higher education* en el motor de búsqueda de Google Académico¹ (Scholar) produce 123.000 resultados entre el año 1990 y el 2000; la misma búsqueda entre los años 2000 y 2010 revela 828.000 resultados, lo que refleja el creciente interés que suscita el tema.
- El sistema de gestión nacional de tesis doctorales TESEO, a cargo del Ministerio de Educación, Cultura y Deporte, también ofrece un motor de búsqueda que permite combinar hasta tres códigos del tesoro de investigación de la UNESCO, con otros campos de consulta habitual. Introduciendo los códigos más generales y próximos a la búsqueda anterior (580106- *Evaluación de alumnos* y 580208- *Educación Superior*) (ver Nota 1), se evidencia una notable diferencia en el número de tesis registradas en el sistema en las dos últimas décadas: frente a los 6 trabajos inscritos en la década de los noventa (cursos 1989/1990 y 1999/2000), entre los cursos 1999/2000 y 2009/2010 dicha cifra asciende a 21. Los primeros cursos de la

¹ Según resultados de la búsqueda efectuada el 5 de octubre de 2014.

segunda década del milenio (desde el 2009/2010 y hasta el 2013/2014) suman 15 trabajos más a la lista, confirmando la tendencia ascendente.

Junto a estas evidencias, la proliferación de jornadas, encuentros y congresos nacionales e internacionales sobre docencia universitaria y evaluación representan una muestra más de la actualidad del tema. También así, la cantidad de institutos, asociaciones o redes profesionales creadas en torno a la inquietud de mejora y desarrollo de la evaluación. A nivel internacional, algunas de las más especializadas en la educación superior son la Association for the Assessment of Learning in Higher Education (AALHE), el Assessment Training Institute (ATI) o la Higher Education Academy (HEA). En el contexto nacional próximo, desempeñan funciones similares el grupo EVALFOR (Evaluación en contextos Formativos) de la Universidad de Cádiz, el Grupo GEM (Grupo Evaluación y Medición) de la Universidad de Valencia o la Red de Evaluación Formativa y Compartida en Docencia Universitaria (A. Pérez et al., 2008; V. M. López, Martínez, & Julián, 2007; V. M. López, Castejón, Sicilia, Navarro & Webb, 2011; V. M. López, 2011). Al importante influjo que ejercen todas ellas, habría que sumar las contribuciones que realizan otras organizaciones dedicadas a la enseñanza superior que, sin estar expresamente comprometidas en la investigación y promoción evaluativa, desarrollan líneas de actuación en torno a ella. La AAHE (American Association of Higher Education), el CFSHE (Centre for the Study of Higher Education) o la QAA (Quality Assurance Agency for Higher Education) son algunas de ellas.

Sin embargo, aunque sean muchos los estudios que se centren en la evaluación del aprendizaje en la universidad, son pocos los que se detienen a analizar la realidad del contexto específico de la formación básica del profesorado (Palacios & López, 2013; Palacios, López, & Barba, 2013). Así se constata al incluir en las búsquedas anteriores la referencia adicional a dichas titulaciones (ver tabla 1):

- Realizando en Google Académico (Scholar) (ver Nota 1) la búsqueda de los términos *assessment*, *higher education* y *teacher education* entre los años 2000 y 2010, el número de resultados desciende de 828.000 a 24.500, el 2,95% de las que se generan al ligar evaluación y educación superior.
- El número de tesis doctorales registradas en TESEO (ver Nota 1) también es menor cuando se añaden condiciones que aproximen el tema a la formación inicial del profesorado. Considerando los códigos del tesauro anteriormente referidos, entre los cursos 1999/2000 y el 2013/2014 figuran 30 trabajos registrados; agregando a la misma búsqueda otros términos que la acercan a las titulaciones de referencia, esta cifra se reduce considerablemente.

Tabla 1

La investigación de la competencia evaluativa en la formación básica de maestras/os. Evidencias indirectas para su cuantificación

Fuente	BÚSQUEDA GENERAL (evaluación y educación superior)	BÚSQUEDA ESPECÍFICA (evaluación, educación superior y formación inicial del profesorado)
GOOGLE SCHOLAR	“Assessment” + “higher education” • 1990-1999. 123.00 resultados. • 2000-2010. 828.000 resultados.	“Assessment” + “higher education” + “teacher education” • 1990-1999. 14.100 resultados. • 2000-2010. 24.500 resultados.
TESEO	“Evaluación de alumnos” (580106) + “Educación superior” (580208) • 89/90-99/00. 6 resultados. • 99/00-09/10. 21 resultados. • 09/10-13/14. 12 resultados. • 99/00-13/14. 33 resultados.	“Evaluación de alumnos” (580106) + “Educación superior” (580208) + ... 99/00-13/14 • “Preparación de profesores” (580203). 1 resultado. • “Pedagogía” (580000). 2 resultados. • “Psicopedagogía” (640400). 0 resultados. • “Sector de la educación” (531204). 0 resultados. • “Educación Básica” (580204). 0 resultados.

La investigación que está siendo introducida responde a esta apreciación y delimita el estudio de la competencia evaluativa docente a las titulaciones vinculadas a la formación básica de maestras/os. De acuerdo a los objetivos expuestos en el próximo apartado, plantea la definición de un modelo teórico para la sistematización de competencia evaluativa docente y un diagnóstico formativo de la misma, contando con la apreciación de estudiantes y profesorado.

En nuestro entorno cercano existen antecedentes que se han interesado por analizar cómo se lleva a cabo la evaluación del aprendizaje en las instituciones dedicadas a la formación de maestras/os o profesorado. Realizados desde perspectivas diferentes, resumimos en orden de sucesión cronológica aquellos que más se aproximan a nuestro planteamiento:

- **M. L. Pérez, Reyes, Palma, & Rafel (2000).** Realizaron un estudio general sobre el tipo de demandas evaluativas que plantea el profesorado universitario y su interpretación y repercusión potencial en el colectivo discente. Inicialmente sus objetivos no estaban enfocados a las titulaciones acotadas; sin embargo, el estudio

se efectuó a partir de una muestra conformada por profesoras/es de la Facultad de Ciencias de la Educación de la Universidad de Girona.

- **Arribas, Carabías, & Monreal (2010).** Desarrollaron un estudio sobre el proceso de enseñanza-aprendizaje en la Escuela de Magisterio de Segovia, en el marco de las titulaciones acotadas. Contando con la perspectiva de docentes, estudiantes de último año de carrera y alumnado egresado, revisaron tres focos de atención: el desarrollo de las competencias profesionales, la metodología y los sistemas de evaluación y calificación. En este último núcleo, publicaron conclusiones sobre el proceso de evaluación, los tipos de examen, la calificación y la participación del alumnado.
- **V. M. López & Palacios (2012).** Llevaron a cabo un estudio para conocer la opinión del alumnado sobre los sistemas empleados en la evaluación de su aprendizaje. Las/os participantes fueron estudiantes del último curso de magisterio (futuras/os maestras/os de diversas especialidades) y de Ciencias de la Actividad Física y el Deporte de las Universidades de Salamanca y Valladolid. Las dimensiones de análisis se corresponden con las abordadas por Arribas et al. (2010) pocos años antes.
- **Hamodi & López (2012).** Plantearon un estudio basado en grupos de discusión con estudiantes de último curso y alumnado egresado de la Escuela de Magisterio de Segovia. Las conclusiones reportadas resumen la experiencia de la evaluación formativa y compartida vivenciada por la muestra durante su proceso de formación básica para el ejercicio docente en Educación Infantil o Primaria.
- **L. F. Martínez et al. (2012).** Efectuaron un estudio para determinar la visión del profesorado y de estudiantes de último curso con respecto al planteamiento de la evaluación. En él, indagaban sobre una serie de cuestiones concretas: la presencia de la evaluación en las programaciones, las capacidades cognitivas priorizadas en las demandas evaluativas, la coherencia entre los elementos del programa y la evaluación en las asignaturas, la relación de esta última con las competencias profesionales previstas, la puesta en práctica de los sistemas evaluativos y su coherencia con respecto a la calificación. La experiencia se contextualizó específicamente en la titulación de Educación Física, impartida desde la Facultad de Formación del profesorado y educación de la Universidad Autónoma de Madrid.
- **Palacios & López (2013) y Palacios et al. (2013).** Dentro de un proyecto más amplio para el análisis de la formación inicial del profesorado de Educación Física, recientemente se han publicado dos estudios contextualizados en la Facultad de

Educación y las Escuelas de Magisterio de la Universidad de Valladolid. Centrados en la delimitación de los sistemas, instrumentos y técnicas de evaluación empleadas por profesorado de dichos centros, sus conclusiones han dan lugar al establecimiento de una tipología docente de tres categorías con rasgos evaluativos diferenciados.

- **Gutiérrez, Pérez & Pérez (2013).** Este estudio describe la visión que sostienen sobre el fenómeno evaluativo docentes y estudiantes actuales y egresadas/os. El trabajo se ha desarrollado desde la Universidad de León, en la titulación vinculada a la formación básica de profesorado de Educación Física. Los resultados comunicados analizan las coincidencias y discrepancias de los colectivos consultados en siete cuestiones específicas vinculadas a la evaluación: cumplimiento de los compromisos evaluativos publicados en el programa, tipo de capacidad cognitiva requerida en las demandas formuladas, frecuencia de comunicación de resultados de aprendizaje, correspondencia de la evaluación con otros componentes curriculares, frecuencia de aplicación de instrumentos y variedad, causas de disfuncionalidad o insatisfacción y desarrollo del proceso de calificación.
- **Ruiz-Gallardo, Ruiz, & Ureña (2013).** De forma reciente han publicado las conclusiones de un estudio desarrollado con estudiantes y profesorado de la diplomatura de magisterio (especialidad en Educación Física) y de Licenciatura en Ciencias de la Actividad Física y el Deporte de tres universidades españolas (Universidad de Castilla-La Mancha, Universidad de Murcia y Universidad Católica San Antonio de Murcia). En él se ha contrastado el punto de vista docente y discente con respecto a cuatro variables asociadas a la praxis evaluativa: programación de la evaluación, instrumentos puestos al servicio de la misma, las capacidades cognitivas captadas y creencias sobre tópicos.
- **Villa, Campo, Arranz, Villa, & García (2013).** Dentro de los trabajos más próximos en el contexto formativo y el tiempo, este estudio ha llevado a cabo la descripción y el contraste de la visión que sostienen profesorado y gestoras/es académicos de magisterio sobre el aprendizaje basado en competencias. Las conclusiones comunicadas hacen referencia a cinco factores asociados al planteamiento pedagógico subyace, entre los que nos resultan de particular interés los dos últimos: planificación de competencias, gestión pedagógica, coordinación docente, tutoría y evaluación y revisión y mejora docente.

- **Hortigüela, Pérez, & Abella (2015).** Desarrollado en la Universidad de Burgos, este estudio compara la experiencia de estudiantes de los Grados de Educación Infantil y Primaria en sistemas de evaluación tradicionales y formativos. De todos los enumerados, es el único que compara los efectos de dispositivos evaluativos diferentes para una misma selección de asignaturas.

En el contexto internacional también se han realizado aproximaciones a la evaluación del aprendizaje practicada en la formación básica de maestras/os. Entre ellas, se encuentra el trabajo de Smitter (2006), que indagó sobre el punto de vista de alumnado avanzado en instituciones dedicadas a la formación del profesorado. C. Brew et al. (2009) se centraron específicamente en las fórmulas participativas de evaluación y calificación, comparando el punto de vista de estudiantes y profesorado al respecto. Aunque desde una perspectiva diferente, también se han aproximado al tema Stiggins (1998) y Stiggins y Conklin (1988). Ambos analizaron la presencia que tienen los módulos o asignaturas asociados al entrenamiento o capacitación evaluativa en los programas de formación del profesorado de distintos estados de los Estados Unidos.

La enumeración de antecedentes realizada sugiere la configuración de una incipiente línea de investigación nacional durante los últimos años y a ella se suma el presente trabajo. Aunque con matices y perspectivas diferentes, su desarrollo trata de mejorar las prácticas evaluativas que tienen lugar en el espacio formativo acotado (formación básica de maestras/os y profesorado), como enclave estratégico para la difusión de los avances obtenidos a los niveles educativos inferiores. Supone un paso adelante con respecto a la investigación clásica realizada en materia evaluativa, caracterizada por prestar atención prioritaria a las herramientas y procedimientos empleados (M. Martínez, García, & Porto, 2012; Tejedor, 1999; Verdejo, 2008). Frente a ese enfoque instrumental, la mayor parte de los antecedentes próximos reseñados y nuestro trabajo ofrecen un planteamiento amplio de la evaluación, considerando su potencialidad como recurso para la promoción del aprendizaje, y no sólo su medición. Ello lleva a indagar sobre toda una serie de factores vinculados al diseño y la puesta en práctica de la evaluación, sin exclusión de los instrumentos (funciones desempeñadas, agentes implicados en su desarrollo, explicitación de bases y condiciones evaluativas, comunicación de resultados y procesos de retroalimentación, etc.).

Dentro de esa misma línea, la presente investigación reporta dos contribuciones adicionales respecto a los trabajos que le preceden:

- La más importante de ellas es que incluye un marco teórico de referencia para conceptualizar la evaluación del aprendizaje de calidad en el contexto universitario

acotado y, desde ella, la competencia evaluativa docente. De acuerdo con Buendía y López (1999) o Porto (2006), en las universidades no existe una propuesta o visión compartida sobre la calidad y el alcance de la evaluación del aprendizaje discente. Nuestro trabajo incorpora en este sentido, dos propuestas interdependientes que establecen un sistema ordenado de indicadores para conceptualizar la evaluación del aprendizaje de calidad y la competencia docente requerida, respectivamente.

- Otro rasgo que caracteriza la aportación del presente trabajo es que trasciende el ámbito específico de la Educación Física, el más representado dentro de los antecedentes identificados (Gutiérrez et al., 2013; V. M. López & Palacios, 2012; L. F. Martínez et al., 2012; Palacios & López, 2013; Palacios et al., 2013; Ruiz-Gallardo et al., 2013). En su lugar, acota el análisis a la formación básica de maestras/os de Educación Infantil o Primaria, y los diversos itinerarios de especialización posibles en la misma, no sólo la Educación Física.

En suma, las razones y motivaciones aducidas nos llevan a ver en la competencia evaluativa docente de estas titulaciones un área de mejora y un foco de atención relevante, pertinente y prioritario. Desde dicha consideración, la investigación se orienta a la consecución de los propósitos declarados en el próximo apartado, en un proyecto que lleva por título ***Competencia evaluativa del profesorado universitario de los Grados de Educación Infantil (EI) y Educación Primaria (EP)***.

2. Objetivos de la investigación

La investigación que está siendo introducida responde a una inquietud general de mejora que es la que la suscita y le da sentido: *favorecer el desarrollo de la competencia evaluativa del profesorado vinculado a la formación básica de maestras/os de Educación Infantil y Primaria, para contribuir a mejorar la calidad formativa en dichas titulaciones y en el sistema educativo en general.*

Para abordar tal propósito, el trabajo se orienta a la resolución de dos preguntas centrales, desglosadas en ocho objetivos:

PREGUNTA 1.

¿Cuáles deberían ser los rasgos que caractericen la competencia evaluativa docente en los grados vinculados a la formación básica de maestras/os de Educación Infantil y Primaria?

Esta primera cuestión representa una inquietud teórica, centrada en la delimitación de referencias que orienten la evaluación del aprendizaje en el contexto acotado y la competencia docente requerida a tal efecto. Su desarrollo se articula en cuatro objetivos, resueltos progresivamente en el marco teórico de la investigación.

● **OBJETIVO A. Formular una propuesta revisada de las principales condiciones y características que debe reunir la evaluación del aprendizaje en el contexto de la formación básica de maestras/os.**

Este primer objetivo propone delimitar una teoría de la evaluación que establezca los rasgos y requisitos definitorios de las prácticas de calidad en este terreno. Considerando que la evaluación está naturalmente sujeta a la divergencia de opiniones e interpretaciones, el desarrollo de la competencia evaluativa docente debe partir, necesariamente, de la reflexión sobre el tipo de práctica hacia la que se desea tender.

Admitido que no existe una única alternativa, la respuesta a este primer objetivo se ofrece en la definición del *Modelo ITC (Integrado Teórico Complejo) para la evaluación del aprendizaje* (capítulo cuarto, marco teórico). Aunque el mismo ofrezca una visión parcial y particular de la praxis evaluativa, su construcción se fundamenta en la consideración de referencias diversas, todas ellas relevantes y de máxima actualidad. El resultado es un sistema ordenado de subprocesos, criterios e indicadores, que delimitan la visión de la evaluación de la que parte la presente investigación, y a la que se orientan sus contribuciones con respecto a la competencia evaluativa docente.

● **OBJETIVO B. Contribuir a la superación de la ambigüedad terminológica definiendo una clasificación revisada de tipologías o modalidades evaluativas que ofrezca un marco común de referencia.**

Este segundo propósito representa, en realidad, una exigencia operativa para poder contribuir al objetivo anterior y a todos los que le siguen. La prolijidad léxica y terminológica que rodea a la evaluación genera confusión y ambigüedad,

dificultando la comunicación, el intercambio y la transferencia. Para ponerle remedio, el capítulo segundo del marco teórico establece las bases conceptuales necesarias para apoyar esta investigación y cualquier otra que comparta sus lineamientos de partida. En concreto, enfrenta el reto de revisar y reordenar las clásicas tipologías evaluativas, para establecer un marco de referencia interno unívoco y actualizado.

● **OBJETIVO C. Diseñar un modelo de competencia evaluativa para orientar procesos de formación y evaluación en el profesorado universitario dedicado a la formación básica de maestras/os.**

El tercer propósito apunta a la definición de un modelo que sistematice por medio de dimensiones, indicadores y descriptores la competencia evaluativa deseable en el profesorado dedicado a la formación básica de maestras/os. Presentado en el capítulo quinto del marco teórico, dicho modelo ofrece una visión parcial de la praxis evaluativa, erigida sobre la teoría de la evaluación declarada (objetivo A) y la consulta de otras fuentes actuales y relevantes. La finalidad prevista para el mismo es aportar las referencias necesarias para emprender procesos de mejora, ya sea por medio de la formación y/o de la evaluación.

● **OBJETIVO D. Identificar los tópicos evaluativos presentes en las escalas de evaluación de la calidad docente administradas al alumnado en la formación básica de maestras/os (trabajo de campo intermedio).**

Este cuarto objetivo desempeña una función subsidiaria con respecto al anterior. Plantea una indagación empírica para proporcionar referencias actuales y reales en las que apoyar la construcción del modelo de competencia evaluativa docente (objetivo C).

Tras la evaluación de la docencia que se efectúe en cada caso, se encuentra latente el modelo de calidad teórico al que se desea tender (Ramsden, 1991; Rizo, 1999; Tejedor, 2003). Partiendo de tal premisa, este cuarto objetivo plantea el análisis de los tópicos evaluativos considerados en los procesos de evaluación de la calidad docente, en instituciones nacionales dedicadas a la formación básica de maestras/os. Articula para ello, un trabajo de campo intermedio sobre las escalas y cuestionarios administrados en 63 universidades españolas que imparten dichas titulaciones. El desarrollo y las conclusiones del mismo se resuelven en el capítulo quinto del marco teórico.

PREGUNTA 2.

¿Cuáles son los rasgos que caracterizan en la práctica la competencia evaluativa docente en los grados vinculados a la formación básica de maestras/os de Educación Infantil y Primaria?

Esta segunda pregunta conecta con el marco empírico de la investigación. Propone averiguar cómo tiene lugar en la práctica la competencia evaluativa, adoptando como referencia comparativa el modelo teórico delimitado en respuesta a la primera pregunta (objetivo D). Dicha exploración parte del testimonio de docentes y estudiantes y aterriza en el contexto institucional de la Universidad de Deusto (UD) y del centro adscrito Begoñako Andra Mari (BAM), ambos sitios en Bilbao (Vizcaya) y vinculados a la formación básica de maestras/os.

La respuesta a la segunda pregunta guía se ha concretado en cuatro nuevos propósitos que, a su vez, se desglosan en sub-propósitos u objetivos más precisos.

● **OBJETIVO E. Realizar una autoevaluación de la competencia evaluativa del profesorado dedicado a la formación básica de maestras/os de Educación Infantil y/o Primaria en el contexto institucional de la UD y BAM.**

El desarrollo de este objetivo se concreta en tres sub-propósitos interrelacionados:

- ✓ **OBJETIVO E1.** Identificar los factores e indicadores de la evaluación del aprendizaje en los que mayor y menor dificultad expresa el profesorado, atendiendo al nivel de relevancia y realización que les atribuyen.
- ✓ **OBJETIVO E2.** Identificar relaciones entre distintas variables académicas, socio-demográficas o profesionales y la competencia evaluativa autopercebida.
- ✓ **OBJETIVO E3.** Identificar las principales fortalezas, debilidades y dificultades para el ejercicio de la evaluación del aprendizaje desde el punto de vista docente.

● **OBJETIVO F. Realizar un diagnóstico de la competencia evaluativa del profesorado dedicado a la formación básica de maestras/os de Educación Infantil y/o Primaria en la UD y BAM desde la perspectiva de las/os estudiantes.**

El sexto objetivo también se concreta en tres sub-propósitos complementarios:

- ✓ **OBJETIVO F1.** Identificar los factores e indicadores de la evaluación del aprendizaje en los que mayor y menor dificultad acusa el alumnado, atendiendo al nivel de relevancia y realización que les atribuyen.
- ✓ **OBJETIVO F2.** Identificar relaciones entre distintas variables académicas, profesionales o socio-demográficas y la competencia evaluativa percibida por el alumnado.
- ✓ **OBJETIVO F3.** Identificar las principales fortalezas y debilidades en la evaluación del aprendizaje desde el punto de vista discente.
- **OBJETIVO G.** Contrastar el punto de vista que manifiestan docentes y discentes con respecto a la competencia evaluativa del profesorado vinculado a la formación básica de maestras/os en la UD y BAM, indagando tanto en la perspectiva teórica (importancia) como en la práctica (realización).
- **OBJETIVO H.** Describir las experiencias y demandas que plantea el profesorado con respecto a la formación permanente vinculada a la evaluación del aprendizaje.

El último propósito conecta con una de las cuestiones que con mayor frecuencia señala la literatura, tanto para explicar las causas de las disfuncionalidades evaluativas actuales, como para proponer vías para su resolución. Ahonda en la formación permanente que realiza el profesorado para el ejercicio de la evaluación y en las demandas y necesidades que manifiesta al respecto. Para ello, se han definido dos sub-propsitos:

- ✓ **OBJETIVO H1.** Describir los principales rasgos de las actividades formativas realizadas por el profesorado en los últimos años con relación a la evaluación del aprendizaje.
- ✓ **OBJETIVO H2.** Identificar las principales necesidades formativas manifestadas por el profesorado con relación al desarrollo de su competencia evaluativa.

3. Proceso de investigación

Para ofrecer respuesta a las dos preguntas centrales y a los ocho propósitos que las desglosan, la investigación ha transcurrido en un proceso ordenado de siete fases, desarrollado entre los cursos 2008/2009 y 2013/2014. La secuencia es el resultado de la adecuación de la propuesta operativa que realizan Sabariego y Bisquerra (2009b) para guiar procesos de investigación y se concreta según representa la figura 1:

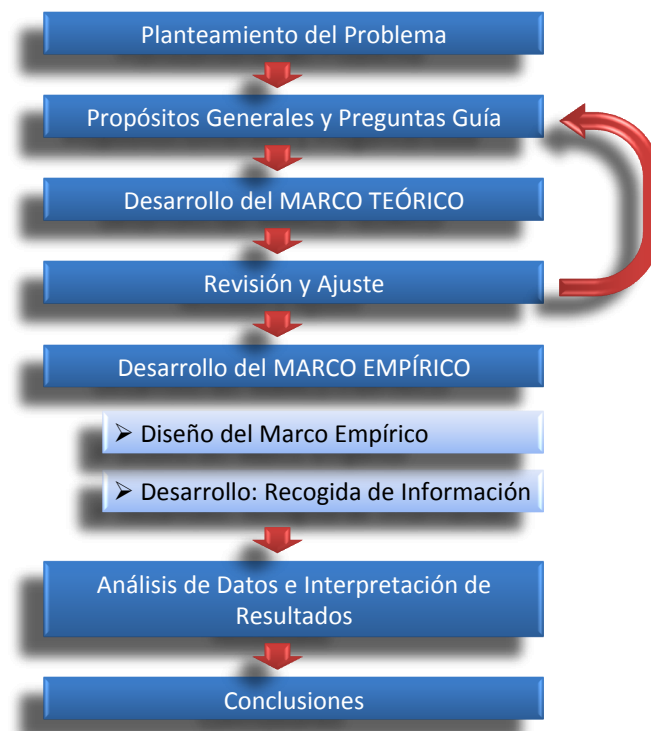


Figura 1. Fases recorridas en el proceso de investigación.

- El *planteamiento del problema* representa la concreción formal de la génesis de la investigación. Supone convertir las motivaciones e inquietudes personales y profesionales de partida en un proyecto formal de investigación. Esta fase tuvo lugar durante el curso 2008/2009, impulsada por la emergencia de un nuevo paradigma formativo, y acelerada por la asistencia a jornadas y congresos y la realización de lecturas específicas.
- La culminación formal de dicho proyecto condujo a la *formulación de las preguntas guía de la investigación y el establecimiento de sus principales propósitos*.

- Las primeras conclusiones obtenidas en la construcción del marco teórico dieron lugar a una cuarta fase de *revisión y ajuste de las preguntas guía y los objetivos de la investigación*, hasta llegar a la formulación presentada en el anterior apartado.
- El *desarrollo del marco empírico* representa la quinta fase en esta secuencia. Encierra en sí dos subfases relevantes: el diseño del trabajo de campo, que terminó de perfilarse a comienzos del curso 2011/2012, y la aplicación o implementación del mismo (segundo cuatrimestre del curso 2011/2012 y primer cuatrimestre de 2012/2013).
- El *análisis e interpretación de los resultados* se ha llevado a cabo ligado al desarrollo del marco empírico. Concluida la primera fase de recogida de información en el tramo final del curso 2011/2012, se hizo un primer registro y los correspondientes análisis tentativos. Una vez finalizada la fase segunda, se procedió a la configuración definitiva de las bases de datos y a la ejecución de los análisis previstos sobre las mismas.
- La fase final de elaboración de *conclusiones* ha tenido lugar entre los cursos 2012/2013 y 2013/2014, recta final del trabajo presentado.

4. Estructura del trabajo

El trabajo está estructurado en cuatro grandes apartados y diez capítulos para ofrecer una memoria completa y ordenada de la investigación realizada:

- El primero de ellos es el **planteamiento general de la investigación** que está concluyendo y que tiene por objeto aproximar al lector/a al contenido que se desarrolla en los próximos capítulos.
- El segundo apartado es el **marco teórico de la investigación**, que establece las bases sobre las que se erige el trabajo de campo. Conformada por seis capítulos interrelacionados, su desarrollo responde a la primera pregunta guía de la investigación y a los cuatro propósitos que la desgranar. Como recoge la figura 2, cada uno de ellos aporta un elemento para la construcción de un modelo teórico que sistematice la competencia evaluativa del profesorado en el ámbito formativo acotado:



Figura 2. Teoría de la competencia evaluativa docente: Estructura interna del marco teórico.

- ✓ El *capítulo primero* se encarga de contextualizar la competencia evaluativa docente en el marco natural en el que se manifiesta. Realiza para ello una breve descripción de la situación actual de la docencia universitaria y de las competencias profesionales que a la misma se atribuyen. En ese contexto, justifica la elección de la competencia evaluativa docente como núcleo prioritario y estratégico de intervención para la mejora de la calidad formativa, en la formación básica de maestras/os, y en el sistema educativo general.
- ✓ El *segundo capítulo* construye un marco conceptual de referencia sobre la evaluación del aprendizaje, en respuesta a las necesidades prácticas que dicta su constatada polisemia. Establecidas las bases para favorecer la interpretación unívoca del proceso evaluativo y sus funciones, incluye una propuesta revisada de las principales modalidades evaluativas para satisfacer el segundo objetivo previsto (B).
- ✓ Establecido el marco de referencia en los dos capítulos anteriores, el *tercer capítulo* recoge el estado de la cuestión. Esto es, analiza cómo tiene lugar

en la práctica la evaluación del aprendizaje en la universidad, según se desprende de la investigación y la literatura específica. Además de los rasgos que mejor caracterizan la praxis evaluativa, el capítulo ofrece una visión sintética de los factores profundos y estructurales que los condicionan, así como de las vías de actuación habitualmente sugeridas para la mejora.

- ✓ El *capítulo cuarto* define aquello que en la investigación en curso se considera una práctica evaluativa de calidad, es decir, establece la teoría de la evaluación del aprendizaje asumida. El resultado es el *Modelo ITC*, que ofrece una visión particular sobre la praxis evaluativa deseable en el contexto formativo acotado. Su definición se fundamenta en la consulta de referencias actuales y relevantes y responde al primer objetivo de la investigación (A).
- ✓ El *quinto capítulo* del marco teórico, aterriza de lleno en la competencia evaluativa del profesorado de los Grados de Educación Infantil y Primaria. Considerando la teoría de la evaluación asumida como referencia (modelo ITC), el capítulo establece una propuesta detallada de la competencia evaluativa que precisa el profesorado para su materialización práctica. En respuesta al tercer propósito (C), configura un sistema ordenado de dimensiones, indicadores y descriptores, para servir de referencia en procesos de formación y/o evaluación de dicha competencia. Nuevamente, el modelo resultante ofrece una visión parcial de la actuación docente en materia evaluativa, aunque parte de un trabajo previo de revisión que lo respalda. Se apoya en la síntesis de los principales antecedentes en la sistematización de la competencia evaluativa, así como en el análisis de herramientas de probada validez y fiabilidad empleadas para su exploración empírica. En correspondencia con el propósito cuarto (D), también incorpora los principales tópicos evaluativos que recogen las escalas aplicadas para la evaluación de la calidad docente en las universidades españolas vinculadas a la formación básica de maestras/os.
- ✓ El *sexto y último capítulo* plantea el balance de las conclusiones extraídas en el marco teórico de la investigación, en respuesta a la primera pregunta guía. Recorre para ello los cuatro propósitos asociados a ella, señalando en cada caso las contribuciones más relevantes elaboradas desde la revisión y el contraste de la investigación y la literatura.

- El tercer apartado del trabajo es el dedicado al **marco empírico**. En él se concreta la recogida y el análisis de información efectuada para responder a la segunda

pregunta guía de la investigación y a los cuatro objetivos vinculados a ella. Su desarrollo tiene lugar en tres capítulos representados en la figura 3:

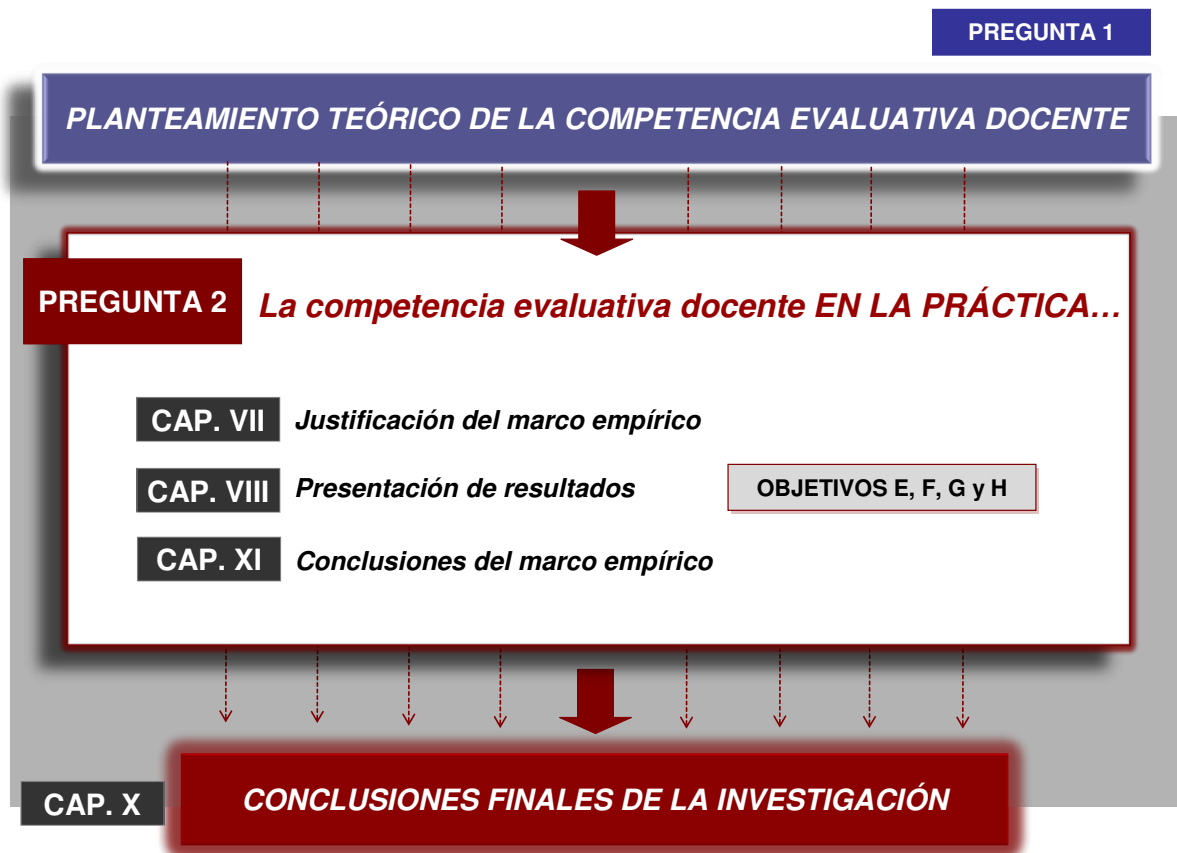


Figura 3. La competencia evaluativa docente en la práctica: estructura del marco empírico de la investigación y conclusiones.

- ✓ El *capítulo séptimo* presenta y justifica el diseño del estudio empírico efectuado a partir del modelo de competencia evaluativa docente definido. En él se detallan la población y la configuración de la muestra, las variables consideradas, los instrumentos y los procedimientos de la recogida y la previsión de los análisis a efectuar a partir de la misma.
- ✓ El *capítulo octavo* contiene los resultados obtenidos en cada uno de los cuatro objetivos asociados al marco empírico de la investigación. Presenta la interpretación que se desprende de los análisis efectuados, antes de proceder a su discusión.
- ✓ El *noveno capítulo* hace balance de las principales conclusiones obtenidas en respuesta a la segunda pregunta guía de la investigación, abordada en el marco empírico. Para ello, contrasta los resultados evidenciados en la

muestra analizada con los hallazgos y las contribuciones que se desprenden de la investigación y la literatura específica. En definitiva, realiza un proceso intermedio de discusión de los resultados relativos a los cuatro últimos objetivos, antes de presentar las conclusiones generales de la investigación en el capítulo final. Plantea así una reflexión sobre los principales acuerdos y discrepancias que se producen al investigar sobre la competencia evaluativa del profesorado, prioritariamente en las titulaciones asociadas a la formación básica de maestras/os.

- El apartado final del trabajo se reserva a la presentación de sus principales **conclusiones**. Se desarrolla en un único capítulo que resume las principales conclusiones extraídas con respecto a las dos preguntas guía de la investigación y a los ocho objetivos dispuestos para su indagación. También incluye una valoración de las contribuciones y limitaciones de la investigación realizada, así como la propuesta de líneas y sugerencias de mejora para continuar trabajando sobre ella.

El recorrido por los apartados anteriores culmina con la exposición de las referencias bibliográficas efectuadas en su desarrollo, y organizadas de acuerdo con la sexta edición del manual de publicaciones de la American Psychological Association (APA, 2010). La memoria se completa con una colección de 30 anexos, mencionados a lo largo del trabajo y presentados en formato electrónico en el CD adjunto.

MARCO TEÓRICO

CAPÍTULO 1.

LA EVALUACIÓN DEL APRENDIZAJE: COMPETENCIA DOCENTE CLAVE EN LA ENSEÑANZA UNIVERSITARIA

Este primer capítulo sitúa la competencia evaluativa docente en el marco de las competencias específicas atribuidas actualmente al profesorado universitario. Destacada clásicamente por su complejidad y relevancia, en las próximas páginas se presentan las razones que han motivado su elección como principal objeto de estudio en la presente investigación.

Para satisfacer dicho objetivo, el capítulo se estructura en cuatro apartados que desarrollan la siguiente línea argumental:

- El primer apartado plantea una breve descripción del contexto formativo específico en el que se enmarca la docencia universitaria desde finales de la década de los noventa. Determinado en gran medida por el planteamiento y la progresiva consolidación del Espacio Europeo de Educación Superior (EEES), en estas páginas iniciales se realiza una breve revisión de los roles atribuidos al profesorado en el nuevo marco de actuación y de las principales dificultades que enfrenta en su desempeño.
- El segundo apartado aborda la delimitación de los perfiles profesionales que ha tenido lugar para el profesorado universitario a raíz del proceso de convergencia. Efectuada una primera descripción de los rasgos generales que acompañan a los mismos, el apartado concluye con la presentación del modelo de competencias asumido como referencia en la investigación en curso: los estándares de calidad y excelencia docente de la Universidad de Deusto.
- En el marco de dicha propuesta, el apartado tercero aterriza de lleno en la evaluación del aprendizaje, competencia específica docente a la que se orienta nuestro trabajo. Efectuada una primera delimitación conceptual sobre su sentido y significado, se exponen en el mismo las razones que justifican su elección como competencia destacada y área de intervención prioritaria.

- El capítulo concluye con una síntesis que rescata las principales conclusiones extraídas de su desarrollo.

1. La docencia universitaria en el marco del EEES

Como indica Zabalza (2012) “ninguna actividad humana es *context free*” (p. 2), incluidas las acciones profesionales, que surgen en un contexto determinado que las envuelve y condiciona. En el caso de la enseñanza universitaria, ese contexto está mediatizado actualmente por el Espacio Europeo de Educación Superior (EEES), proyecto que arranca a finales de la década pasada con las Declaraciones de La Sorbona (1998) y Bolonia (1999), y que plantea la configuración progresiva de un panorama formativo con rasgos propios y diferenciados.

La universidad es una institución medieval, enfrentada permanentemente al reto de ofrecer respuesta a las demandas y necesidades de la sociedad, sin dejar de lado sus valores más preciados (Parcerisa, 2010). Ese cometido, y la búsqueda incesante de calidad en la que desemboca, hace del cambio una constante en su evolución e historia (De Vries, 2005; Hannan & Silver, 2005; Michavila, 2005). Así, en el contexto español, la cronología del cambio en la enseñanza universitaria contempla 20 hitos relevantes entre los años 1808 y 2001 (Valcárcel, 2003). El proceso de convergencia al EEES surge en el tramo final de dicha evolución, por la confluencia de una serie de factores de orden político, económico, social, científico y tecnológico (Cid-Sabucedo, Pérez-Abellás, & Zabalza, 2009; Rué, 2007, 2009a; Villa, 2004).

La universidad tiene que afrontar con urgencia un debate sereno y riguroso sobre cuestiones tan nucleares como su función social, su sistema de gobierno, su carácter de institución pública, el papel del estudiantado en la institución o el modelo de profesorado. Son cuestiones que el proceso hacia el EEES ha puesto de evidencia pero que se deberían debatir con o sin este espacio. (Parcerisa, 2010, p. 19).

La aproximación al proyecto del área europea de la educación superior da lugar a un nuevo ciclo de renovación y cambio (ver figura 4). Concluida en 2010 una primera etapa de

construcción y consolidación, actualmente la universidad se encuentra inmersa en una segunda fase de revisión y consolidación, que fija el 2020 como nuevo plazo para la superación de los retos pendientes (Comas, 2013).

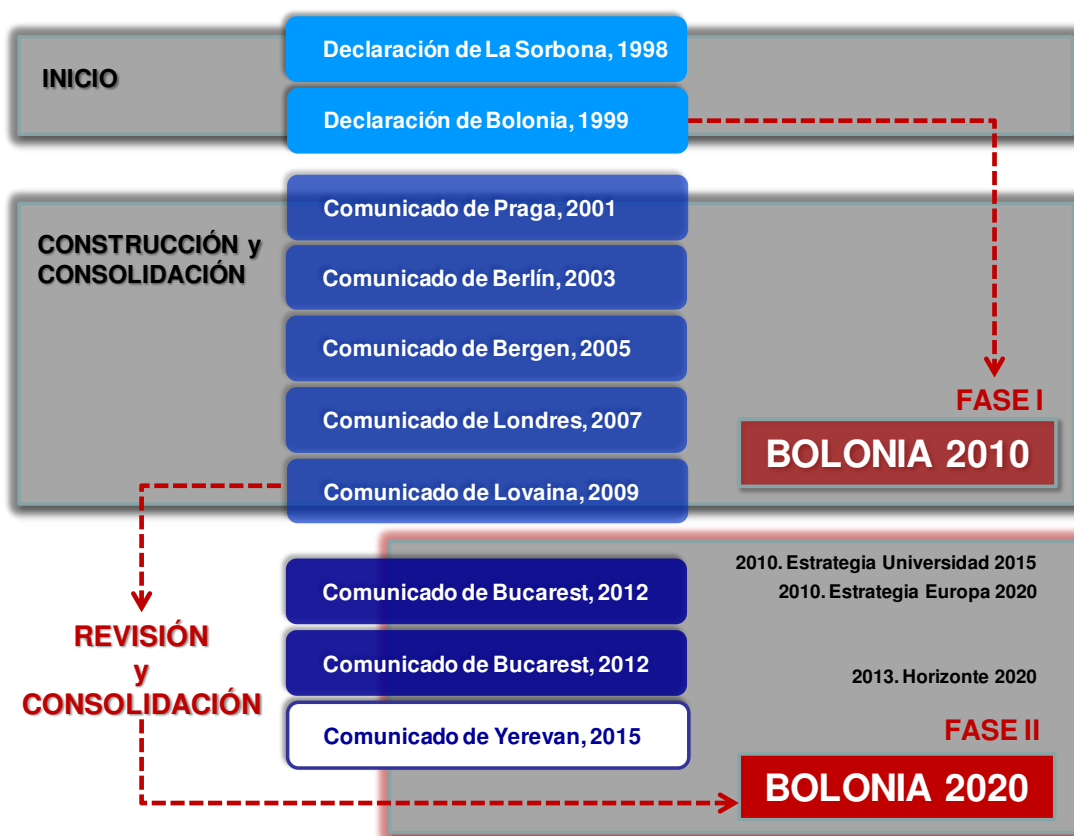


Figura 4. El proceso de Bolonia: cronograma básico. Elaboración propia a partir de Comas (2013), European Higher Education Area (EHEA, 2010) y Estrategia Europa 2020 (Comisión Europea, 2010).

La política de la convergencia “tiene entre sus propósitos fundamentales fortalecer las universidades como centros de desarrollo del conocimiento, de la investigación y del aprendizaje en situación competitiva respecto de las universidades de otros continentes” (Villa, 2004, p. 272). Esa motivación inicial sigue presente en la segunda fase del proceso, aunque los logros efectuados en la primera etapa posibiliten la formulación de aspiraciones de mayor calado. De acuerdo con las decisiones adoptadas en 2009, en la Conferencia de Ministros europeos responsables de la Educación Superior en Lovaina, las principales son:

- Fomentar la igualdad de oportunidades en el acceso y la consecución de los estudios universitarios.
- Promover el aprendizaje a lo largo de la vida.
- Mejorar la empleabilidad de las/os tituladas/os universitarias/os.

- Favorecer una reforma curricular continuada que dé lugar a recorridos educativos de alta calidad.
- Reforzar la unión entre educación, investigación e innovación.
- Avanzar en la internacionalización de la educación superior y en la consolidación del reto de la movilidad entre los países miembros.
- Procurar una financiación prioritariamente pública de la educación superior y estimular la búsqueda de nuevas fuentes de financiación alternativas.
- Desarrollar mecanismos y herramientas para mejorar la transparencia de la información generada en las instituciones de educación superior y favorecer la recogida de datos que favorezcan los procesos de evaluación general.

Para favorecer la realización efectiva de tales propósitos, el proceso de convergencia proyecta una profunda modificación de la educación superior en fondo y forma; en consecuencia, a menudo es apuntado como el mayor cambio realizado en la historia de la universidad (L. C. Contreras, 2007).

Cuando estamos trabajando en estos cambios en profundidad comprendemos que la convergencia se produce no sólo en Europa o con un marco de referencia europeo (EEES); sino también, la convergencia se está produciendo hacia dentro de nuestras instituciones, afectando a todos sus miembros y a todos los estamentos de la institución y sus prácticas. Y ahí está nuestro mayor reto: aprender desde las acciones de cambio y su impacto, e institucionalizar la innovación. (Cebrián, 2008, p. 22).

En la última década se ha generado una extensa literatura sobre los retos y cambios que supone la incorporación al EEES y las oportunidades y riesgos detectados en el camino. No es objeto de la presente investigación insistir o profundizar en ellos, pero entender la docencia universitaria, exige reparar en las circunstancias que rodean la enseñanza superior en los últimos años. La revisión e integración de varias de las aportaciones que a este respecto se han realizado, nos lleva a concluir que el proceso de convergencia plantea cambios a dos niveles complementarios e interdependientes (ver tabla 2):

Tabla 2

Principales cambios pedagógicos y estructurales vinculados al proceso de Bolonia

CAMBIOS ESTRUCTURALES	CAMBIOS PEDAGÓGICOS
a. Configuración de la Europa del Conocimiento sobre dos pilares relevantes: el EEES y el EEI (Espacio Europeo de Investigación) para la movilidad de personal y alumnado. - Nueva estructura de los estudios superiores (grado y postgrado). - Sistema ECTS (European Credit Transfer System). - SET (Suplemento Europeo al Título). - Desarrollo de marcos legales para la promoción de la investigación, la transferencia y la innovación internacional^a. b. Propuesta y verificación de los grados y postgrados como proyecto formativo de la institución que los acoge. c. Acreditación de la calidad en las titulaciones y de la docencia.	a. La formación basada en el aprendizaje. b. La formación basada en competencias. c. La renovación metodológica en el nuevo contexto universitario. - Desarrollo y evaluación de competencias. - El alumnado y su actividad como núcleo central de la formación. d. El nuevo rol discente, más activo y autónomo. e. El nuevo rol docente.
NUEVA ESTRUCTURA FORMATIVA	NUEVO PARADIGMA FORMATIVO

Nota: Síntesis realizada a partir de varias fuentes (Arénaga, 2013; Benito & Cruz, 2005; Bejarano, 2008; Caurcel & Morales, 2008; Cebrián, 2008; Comas, 2013; L. C. Contreras, 2007; Hernández, 2005; J. M. Esteve & Gisbert, 2011; Goñi, 2005; Mas, 2011; Mateo, 2006b, 2008; Mateo & Vlachopoulos, 2013; Ministerio de Educación, 2010; Parcerisa, 2006, 2010; Ramos, 2008; Rué, 2007; J. M. Ruiz, 2010; Villa & Poblete, 2011; Zabalza, 2006, 2011).

^a En el contexto español, destaca por su relevancia la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación.

Especialmente relevantes en la fase inicial del proceso, los cambios estructurales surgen para favorecer la transferencia y la transparencia (Valcárcel, 2003). El deseo de los países miembros de disolver sus fronteras formativas, y potenciar con ello la movilidad de investigadoras/es, docentes y estudiantes, genera la necesidad de acordar una estructura y una métrica común que posibilite el reconocimiento mutuo. Salvadas dichas condiciones por medio de procesos de verificación y acreditación, se dota a las instituciones formativas de autonomía para proponer sus propias titulaciones de grado y postgrado, favoreciendo el incremento de la competitividad general del sistema universitario. La segunda fase del proceso insiste particularmente en dicha competitividad, reforzando la unión de educación, investigación e innovación y su aplicación a la promoción económica, social y cultural. Así se constata en la Estrategia Universidad 2015 (Ministerio de Educación, 2010).

Aunque su incidencia potencial en la renovación sea mayor (Fullan, 2007), los cambios pedagógicos encuentran un desarrollo menos evidente hasta la fecha (Michavila, 2011; F. Pérez, Serrano, Pastor, Hernández, & Zaera, 2012; Vázquez, 2011). Están

llamados a configurar un paradigma formativo compartido para la educación superior, caracterizado por centrar la atención en el aprendizaje discente, y no en la enseñanza, como tradicionalmente venía haciendo (B. Álvarez et al., 2007; Caurcel & Morales, 2008; Feixas, 2008; Morales, 2006a, 2008c; Torre, 2004). Por influjo de los cambios estructurales, la definición de dicho aprendizaje se realiza en términos de competencia, favoreciendo la equiparación de los estudios y una mejor empleabilidad para las/os egresadas/os (Viso, 2010). Todo ello suscita la exploración y aplicación de nuevas metodologías, más centradas en la actividad discente, y más acordes a la naturaleza y los rasgos específicos de las competencias (Alfaro, 2006; Cano & Ion, 2012; Palomares, 2011; Zabalza, 2011).

En este escenario se produce una redefinición forzosa del papel que juegan estudiantes y profesorado, como principales agentes involucrados en la acción formativa:

- Para el desarrollo efectivo de las competencias previstas en el perfil profesional de la titulación, se requiere del alumnado una mayor implicación y participación en el proceso formativo, de modo que desempeñe el papel activo que exige el aprendizaje desde la óptica constructivista (Blanco & Alba, 2009; L. Prieto, 2004; Rué, 2007; Zabalza, 2002). Más aún. Para afrontar los retos de una sociedad en expansión las/os estudiantes deben desarrollar la capacidad de aprender a lo largo de la vida (Hernández, 2005; Hernández, Martínez, Martínez, & Monroy, 2009; Villardón, Yániz, Elxpuru, & Achurra, 2013).
- En cuanto al rol docente, el planteamiento del EEES da lugar a una nueva visión de la enseñanza universitaria que afecta a sus roles, prioridades y modos de actuación. La definición sistemática de la calidad docente representa una inquietud permanente en la historia de la universidad, pero ha conducido a interpretaciones divergentes y discrepantes de forma recurrente (Cabrera & La Nasa, 2005; J. M. Escudero, 2006; Gros & Romañá, 2004; García-Valcárcel, 1992; Jackson, 2006; Murillo, 2008; L. Prieto, 2007). En este contexto, la visión revisada de la profesión que subyace a la convergencia se instala como referencia general y compartida de lo que debe ser la enseñanza universitaria. El reto de aproximar a esa visión las prácticas docentes se introduce en la primera fase del proceso, y aún persiste. De hecho, representa para muchos uno de los factores más determinantes en el cambio de paradigma propuesto (M. Martínez & Viader, 2008; Mas, 2011, 2012; Michavila, 2005, 2011).

Presentado en el próximo apartado, este cambio en el modo de entender la enseñanza universitaria está dando lugar a una fecunda línea de trabajo con dos grandes vertientes: la innovación e investigación sobre el cuadro de competencias presentes en el nuevo perfil profesional (Gros, 2006; Tejada, 2009) y la ampliación y renovación de la oferta

formativa que haga posible su interiorización por parte del profesorado (J. M. Esteve, 2006; Monereo, 2013, 2014a; Montero, 2006).

1.1 El papel del profesorado en el proceso de convergencia.

El EEES plantea un proyecto amplio y ambicioso en el que no es posible prescindir de la participación de ninguno de los agentes involucrados (representantes sociales, responsables políticos y administrativos, instituciones formativas y estudiantes, principalmente). Sin embargo, el profesorado ocupa una posición destacada hasta el punto de “considerarse como una pieza más de la arquitectura que la universidad pone en juego para conseguir una enseñanza de calidad” (Torra et al. 2013, p. 287). Como afirma Zabalza (2011), “ni las leyes, ni los directivos, ni las propias instituciones [...] han conseguido nunca sacar adelante una propuesta innovadora al margen o en contra del profesorado que la ha de aplicar en sus clases” (p. 77). Esta premisa adopta una significación mayor al tratar el proceso de convergencia, donde los cambios estructurales y pedagógicos proyectados modifican de forma directa y sustancial el quehacer docente (Alfaro, 2006; B. Álvarez et al., 2007; M. González & Raposo, 2009; M. Martínez, 2006).

Los apartados presentados a continuación describen los principales espacios de actividad que prevé la universidad de la convergencia para el profesorado, así como las principales dificultades a las que se enfrenta en la superación de los nuevos retos. El objetivo de los mismos es dibujar el marco de referencia general que rodea a la definición de las competencias profesionales docentes en el panorama universitario actual, punto de partida del trabajo de investigación que está siendo presentado.

1.1.1 Espacios de actividad del profesorado en el panorama universitario actual: principales roles y funciones.

Los cambios tecnológicos, políticos y socioeconómicos y el nuevo espacio formativo que se viene configurando desde 2010 afectan a la caracterización general de la figura docente y modifican sus funciones. Para contribuir a la misión encomendada a las instituciones universitarias, tres han sido los roles tradicionalmente atribuidos al profesorado: la docencia, la investigación y la gestión (M. P. García & Maquilón, 2010; Mas & Ruiz, 2007; Sagasta, Bilbatua, Galíndez, & Bikuña, 2006; Torra et. al., 2013; M. Vera,

Salanova, & Martín, 2010). La primera redacción de la LOU (Ley Orgánica de Universidad 6/2001) dejaba constancia de ello al tratar la retribución del personal docente e investigador (Art. 55) y los procesos de habilitación nacional (Art. 58, 59 y 60).

Frente a esa delimitación clásica, los referentes normativos actuales dan cuenta de un primer cambio relevante en las atribuciones del profesorado universitario. A ellas se suma la transferencia, contemplada como cuarta función o espacio de actividad en la modificación de la LOU aprobada en el año 2007 (LOMLOU 4/2007), en la resolución del XXII Convenio Colectivo de ámbito estatal para los centros de educación universitaria e investigación (2012) y en el último borrador del Estatuto del personal docente e investigador de las universidades públicas españolas (Ministerio de Educación, 2011). Este último expone:

La función profesoral no termina con la labor docente. El profesor ha sido, sigue siendo y debe seguir siendo un investigador, un generador de conocimiento y no un mero transmisor. El sistema universitario en su conjunto debe permitir que el profesor compagine ambas tareas, así como otras fundamentales en el seno de la Universidad como la transferencia de conocimiento, la innovación o las tareas de gestión universitaria, mediante el diseño de soluciones legales y de gestión que asuman esa complejidad de la actividad del profesorado universitario, sin perjuicio de la potenciación de alguno de los aspectos por encima de otros, pero sin perder el carácter multifuncional que debe predicarse del profesorado. (p. 2).

“En el contexto de la sociedad del conocimiento hacer visible la utilidad de las instituciones de educación superior para servir a la sociedad constituye una de sus grandes prioridades” (Mateo, 2012, 214). En consecuencia, la transferencia figura entre las principales misiones encomendadas a las instituciones formativas españolas en la Estrategia Universidad 2015 (Ministerio de Educación, 2010), así como en las propuestas de avance articuladas a nivel europeo en la segunda fase del proceso de convergencia. Entre ellas resultan especialmente relevantes la Estrategia Europa, 2020 (Comisión Europea, 2010) y Horizonte 2020 (Comisión Europea, 2013).

Esta cuarta función remite al espacio intelectual y discursivo que sugiere Barnett (2008) para la acción universitaria, aquel destinado a las aportaciones académicas, al discurso social y comunitario. En esencia, la transferencia en el ámbito académico es “el proceso en que el conocimiento generado por la investigación académica se aplica para la

producción de nueva tecnología, capacidades o técnicas que puedan tener valor comercial o son de aplicación práctica” (Ministerio de Educación, 2010, p. 76). Engloba la contribución a la innovación y al desarrollo social y económico, favoreciendo la resolución de los grandes problemas presentes en la actualidad. Así se contempla en la Ley 14/2011 de la Ciencia, la Tecnología y la Innovación ideada para su estimulación.

Según Kindelán (2013) este nuevo cometido ha cobrado especial protagonismo en los últimos años, cuando la emergencia de la tecnología parece restar relevancia a la tradicional labor transmisora ejercida sobre el alumnado. La enumeración de los tres roles clásicos persiste en muchos casos, pero cada vez son más frecuentes las referencias que incluyen la innovación y la transferencia como cometidos del profesorado universitario (Chocarro, Sobrino, & González, 2013; J. M. Ruiz, 2010; Villa, 2008). La idea de servicio a la comunidad está presente en todas ellas, aunque con matices que la aproximan más a la dimensión científica o social. En una integración alternativa de ambas opciones, Zabalza (2002) incorpora a los tres roles tradicionales del profesorado el *bussiness* (negociación de proyectos y convenios, búsqueda de fuentes alternativas de financiación, etc.) y las relaciones institucionales (relaciones con empresas y con otras universidades o instituciones formativas, sociales, etc.).

Pero el proceso de convergencia y los diversos factores que lo impulsan no sólo afectan a la incorporación de esta nueva función, sino que también matizan el sentido y el significado de las ocupaciones clásicas. Asumidas de forma general, la investigación, la docencia y la gestión han ido integrando los ecos de los cambios pedagógicos y estructurales de los últimos años, aunque en distinta medida.

La gestión resulta el menos representativo y relevante de los espacios de actividad del profesorado (Mas, 2011, 2012) y, probablemente, el que en menor profundidad ha cambiado. El artículo 12 del borrador del Estatuto del personal docente e investigador de las universidades públicas españolas (Ministerio de Educación, 2011) incluye dos modalidades de gestión: las tareas de gestión y dirección vinculadas a la docencia, la investigación y la transferencia, y las actividades específicas de dirección, representación y gestión con proyección global.

A pesar de la complejidad organizativa que caracteriza a las universidades, generalmente la gestión representa un espacio de actividad desempeñado de forma temporal, sin cualificación específica y sin renunciar al resto de las funciones (Feixas, 2002). Los cambios estructurales sujetos a la configuración del EEES y, en especial, las exigencias derivadas de los procesos de acreditación han provocado su intensificación en los últimos años; aunque las tecnologías de la información y la comunicación (TIC) hayan aportado

nuevas herramientas para facilitar su desarrollo (F. M. Esteve & Gisbert, 2011), la gestión suscita sensaciones negativas entre el profesorado (Monereo & Domínguez, 2014). Esta situación da lugar a un lamento por la burocratización de la enseñanza (Fonseca, 2010; Zabalza, 2002), así como a la demanda de nuevas figuras o roles *paraacadémicos*, especializados en tareas de gestión (Feixas, 2002).

Estrechamente ligado al espacio de la innovación y la transferencia, la investigación representa otro de los roles clásicos reconocidos para el profesorado universitario desde la perspectiva normativa (borrador Estatuto del personal docente e investigador de las universidades públicas españolas, 2011; Resolución del XXII Convenio Colectivo de ámbito estatal para los centros de educación universitaria e investigación; LOMLOU, 2007). Frente a la docencia y la gestión, que suelen estar claramente reguladas, representa una función más amplia y heterogénea, dando lugar a la coexistencia de niveles de compromiso e implicación muy desiguales dentro de un mismo equipo docente (Mora, 2000).

Como será expuesto en el próximo apartado, la labor investigadora suscita una amplia controversia por las dificultades de compatibilidad o equilibrio que plantea con respecto a la docencia (Hernández, 2002; Tomás, Castro, & Feixas, 2012; Villa, 2008). Aun así, su importancia y pertinencia representa un hecho innegable, y así lo admite el propio profesorado (Monereo & Domínguez, 2014; Vidal & Quintanilla, 2000). De hecho, según Kindelán (2013) “la investigación científica y la transferencia del conocimiento a la sociedad son los puntos neurálgicos que distinguen al profesor universitario de otros docentes en los niveles previos [...]” (p. 38).

Los cambios tecnológicos y sociales de los últimos años y su reflejo en el paradigma formativo de la convergencia han revitalizado la actualidad de la investigación y arrojado ciertos cambios sobre ella. El más evidente de todos ellos es la creación del Espacio Europeo de Investigación (EEI) tras el Consejo Europeo extraordinario de Lisboa (2000).

El concepto de EEI combina los siguientes elementos: un ‘mercado interior’ europeo de la investigación, en el que los investigadores, la tecnología y los conocimientos circulan libremente; la coordinación efectiva a nivel europeo de las actividades, los programas y las políticas de investigación nacionales y regionales; y las iniciativas ejecutadas y financiadas a nivel europeo. (Comisión de las Comunidades Europeas, 2007, p. 2).

Esta apuesta por la investigación y el desarrollo de su competitividad internacional representa una de las grandes apuestas recogidas en la Estrategia Europa 2020 (Comisión Europea, 2010), donde figura estrechamente vinculada a la innovación y la transferencia social y científica. Para el profesorado este impulso se hace especialmente visible en los procesos de acreditación docente, donde el reconocimiento contundente de los méritos de investigación genera una fuente de motivación o revulsivo para su desarrollo (Fonseca, 2010; Hernández, 2014).

Pero además de la intensificación, el rol investigador ha adoptado en los últimos años otras modificaciones relevantes que afectan a su naturaleza y a los medios disponibles para su desarrollo:

- Por una parte, el proceso de convergencia ha contribuido a la creación de un nuevo concepto multilateral de investigación, que traspasa los límites académicos y disciplinares (Kindelán, 2013). Aunque el profesorado sigue considerando que la actividad investigadora es fundamentalmente solitaria (Tomás et al., 2012), se apuesta por la producción interdisciplinar y globalizada, realizada desde equipos de investigación nacionales e internacionales.
- Por otra parte, Mas y Ruiz (2007) afirman que la función investigadora se ha visto modificada por la generalización de nuevos medios y recursos aplicables a su desarrollo. En efecto, la aplicación de las TIC facilita nuevas oportunidades para el diseño, desarrollo y difusión de la labor investigadora (accesibilidad a fuentes bibliográficas y a la investigación reciente, recursos para el procesamiento, registro y almacenamiento de las fuentes consultadas, fácil comunicación de los hallazgos, soportes ágiles para la coordinación y el intercambio, etc.).

La revisión de los principales roles atribuidos al profesorado universitario debe concluir necesariamente con la función docente, aquella orientada a la capacitación profesional del alumnado y la que mejor representa su identidad profesional (Monereo & Domínguez, 2014). Aunque los nuevos medios tecnológicos hayan modificado sustancialmente el rol docente tradicional, este espacio de actividad sigue representando un factor indispensable de la calidad formativa (Paris, 2013; Viso, 2010). Partiendo de esta premisa, buena parte de los cambios pedagógicos que plantea la configuración del EEES inciden directamente en el ejercicio docente.

Sin negar la importancia de la gestión, la investigación y su transferencia a la comunidad social y científica, este trabajo pone el foco de atención en la enseñanza como espacio específico de actividad del profesorado. En consecuencia, el próximo apartado

aborda expresamente las modificaciones que están teniendo lugar sobre la misma desde finales de la década de los noventa.

1.1.1.1 Principales modificaciones en el rol docente.

La renovación estructural y pedagógica que exige la aproximación al área europea de la educación superior reporta cambios importantes sobre los diversos espacios de actividad del profesorado según ha sido expuesto. La función docente, la más visible y prioritaria desde nuestra perspectiva, no ha permanecido ajena a ese proceso. La nueva cultura de los procesos de enseñanza y aprendizaje convierte a las/os docentes en agentes transformadores, anticipadores y críticos, con una fuerte responsabilidad de liderazgo pedagógico en el cambio (A. Medina, Domínguez, & Ribeiro, 2011).

El EEES plantea sobre el rol docente toda una serie de modificaciones, que presentamos resumidas en cuatro grandes cambios:

● Cambio del objetivo u orientación general de la actividad docente: de enseñar, a hacer aprender.

El planteamiento pedagógico que subyace al proceso de convergencia ha vuelto a poner en primera línea de actualidad los enfoques docentes, un objeto de estudio clásico abordado desde diversas perspectivas (Pozo, Scheuer, Mateos, & Pérez, 2006). Siguiendo a Feixas (2002), este constructo hace referencia a la orientación didáctica que imprime el profesorado a su enseñanza y está estrechamente relacionada con sus creencias y concepciones sobre el aprendizaje y la naturaleza de su profesión. Las concepciones constituyen una determinada representación o interpretación de la realidad (Pozo, 2001) y se construyen implícitamente por la exposición reiterada a determinados patrones de actuación (Buendía et al., 1999; Pozo et al., 2006; Reber, 1993), mientras que las creencias son las tendencias o disposiciones que llevan al profesorado a enseñar de un modo determinado (L. Prieto, 2007).

Aunque con evidentes discrepancias terminológicas, la literatura específica da cuenta de dos grandes orientaciones o enfoques docentes con amplio grado de acuerdo: aquel centrado en la enseñanza y el profesorado, y el orientado al aprendizaje y el desarrollo discente (Barr & Tagg, 1995; García-Valcárcel, 1993; B. Gargallo, Suárez, Garfella, & Fernández, 2011; Sánchez, Ros, & Ferraras, 2010; Gow & Kember, 1993; Hernández & Maquilón, 2011; Hernández, Maquilón, Gracia, & Monroy, 2010; Jackson, 2006; Kember & Gow, 1994; Meade, 1997). La

discriminación de estas dos grandes visiones resume propuestas más complejas y pormenorizadas, en las que se delimitan entre tres y seis formas diferentes de entender la enseñanza (Biggs, 2005; Entwistle, 2000; Johnston, 1996; Kember, 1997; Pratt, 1992; Prosser & Trigwell, 1999; Ramsden, 2003; Samuelovicz & Bain, 2001).

En este marco de referencia, el paradigma formativo del EEES demanda para la enseñanza un cambio de dirección que le lleve a centrarse en el aprendizaje discente, y no tanto en el contenido y la docencia (B. Álvarez et al., 2007; Benito & Cruz, 2005; Caurcel & Morales, 2008; Villa, 2004). En palabras de Torre (2004), este movimiento implicaría la ruptura de la clásica “asociación dicotómica profesor-enseña y estudiante-aprende” (p. 11).

La tradicional misión del docente como transmisor de conocimientos ha quedado relegada a un nivel secundario para priorizar fundamentalmente su papel como facilitador del aprendizaje de sus estudiantes. El acceso al conocimiento se puede hacer hoy a través de muy diversas vías [...] pero la facilidad de acceso no supone una ayuda significativa en lo que se refiere a la decodificación, asimilación y aprovechamiento de tal información, ni garantiza la vinculación de dicho material con la práctica profesional. (Zabalza, 2002, p. 110).

El nuevo planteamiento formativo incorpora en sus bases la redefinición de la finalidad docente y de los criterios para valorar su éxito profesional (Morales, 2006a Villa, 2008). Encauzada desde premisas diferentes, su efecto debería hacerse presente desde la fase inicial de planificación de la enseñanza (M. L. Fernández, 2007).

Una consecuencia clara de mirar más directamente al resultado de nuestra acción docente es que deberíamos hablar menos de calidad de la enseñanza y más de la calidad del aprendizaje, o fijarnos en primer lugar en la calidad del aprendizaje (resultados en los alumnos) y a partir de ahí reflexionar sobre la calidad de la enseñanza. (Morales, 2008c, p. 18).

En definitiva, el planteamiento pedagógico que subyace a la convergencia lleva a una modificación sustancial del rol docente. El profesorado deja ser concebido

como experto en la materia encomendado a una labor transmisora, para convertirse en un agente facilitador del aprendizaje (L. C. Contreras, 2007).

● ***Cambio en la ontogénesis o naturaleza del aprendizaje: de la reproducción de contenidos al desarrollo de competencias.***

La reconfiguración del cometido docente tiene lugar junto a la redefinición de la propia naturaleza del aprendizaje. Lejos de interpretarlo como acumulación memorística de contenidos transmitidos, trata de instalarse una visión constructivista basada en el desarrollo de competencias y en la utilización de procesos cognitivos profundos y de orden superior (Biggs, 2005; Caurcel & Morales, 2008; B. Gargallo, Sánchez, Ros, & Ferraras, 2010).

El término competencia ha generado en las últimas décadas un cambio cualitativo en la forma de entender el aprendizaje humano en general. Al referirnos a aprendizaje y desarrollo de competencias, el propio concepto de aprendizaje se amplía y la reflexión sobre el mismo adquiere una nueva dimensión, más sistémica y holística. Este cambio permite concebir los objetivos de la docencia de forma más dinámica y completa. (M. Martínez & Viader, 2008, p. 221).

El enfoque docente basado en el aprendizaje y el alumnado sólo tiene sentido partiendo de dos premisas interrelacionadas que, por el momento, se encuentran poco extendidas en la práctica (Aparicio & Pozo, 2006). La primera de ellas es admitir que el aprendizaje excede los límites del dominio teórico, lo que remite a las competencias como principal referente formativo. La segunda supone adoptar la perspectiva constructivista y abordar el aprendizaje como un objeto no transferible. Concebido como construcción personal por parte de quien lo lleva a cabo (Barth, 2004; Carretero, 1994; Chickering & Gamson, 1987, 1999), la enseñanza pasa a convertirse en un medio o recurso para facilitar dicho proceso externamente (Morales, 2006a).

En este contexto, “el papel del docente universitario consistirá básicamente en crear y facilitar situaciones de aprendizaje que faciliten el desarrollo personal y contribuyan a una sociedad sostenible, democrática y basada en el conocimiento” (Ortega, 2010, p. 322). La participación y actividad discente representa la condición indispensable o el punto de partida (Ahumada, 2002; Parcerisa, 2010), mientras que el profesorado

debe “diseñar las situaciones o escenarios de aprendizaje pertinentes para entrenar en los diferentes niveles de desempeño de las competencias, esto es, para hacer que el alumno haga lo que se espera de él” (Bonsón, 2009, p. 31).

Esta labor de mediación o facilitación es otro de los argumentos que hace del colectivo docente una pieza estratégica en el proceso de convergencia. “Es el profesor quien en última instancia debe modular e ir definiendo operativamente el rol del alumno y sus actitudes hacia el aprendizaje a través de lo que le hace hacer” (Morales, 2008c, pp. 22-23). En consecuencia, su papel es fundamental “para ayudar a los estudiantes a comprender las nuevas posibilidades y contextos de aprendizaje, así como las estrategias más adecuadas para afrontar este proceso de cambio” (L. Prieto, 2004, p. 113).

● ***Cambios en la metodología docente (y las prácticas evaluativas).***

La confluencia de los dos anteriores cambios sitúa al profesorado europeo ante el reto de explorar alternativas didácticas más acordes a la naturaleza de su cometido: favorecer en el alumnado el desarrollo de competencias. Subrayar la importancia del aprendizaje de las/os estudiantes equivale a admitir que, más que nunca, la clave de la formación reside en el profesorado y su labor docente (M. Martínez, 2006).

Ampliando las aportaciones que realizan al respecto varias/os autoras/es del contexto próximo (Alfaro, 2006; L. C. Contreras, 2007; Bejarano, 2008; Bonsón, 2009; Bozu & Canto, 2009; Goñi, 2005; Rué, 2007; Zabalza, 2011) concluimos que la revisión y renovación metodológica está teniendo lugar principalmente sobre los siguientes focos:

- ✓ *La interacción didáctica entre docentes y discentes y el reparto de la responsabilidad.* Tanto desde la perspectiva estructural de los planes de estudio como desde la perspectiva pedagógica, el alumnado pasa a constituirse como el centro de la acción formativa. Ello lleva a demandar una participación discente más activa y decidida y a preservar un mayor espacio para su trabajo autónomo.
- ✓ *La organización del espacio y el tiempo.* El aumento de la participación y la autonomía del alumnado provoca que el trabajo deje de estar circunscrito al espacio y horario del aula. De acuerdo con el sistema ECTS, las/os estudiantes deben desarrollar buena parte de su trabajo sin regulación docente directa. Incluso el que desarrollan dentro del ámbito lectivo,

transcurre en condiciones novedosas para responder a las exigencias del aprendizaje de competencias. Junto a la clásica modalidad didáctica magistral, se organizan otro tipo de sesiones para favorecer la aplicación y transferencia y se incrementan los espacios para tutoría.

- ✓ *Las tareas y actividades de aprendizaje (y evaluación).* Definir los objetivos de aprendizaje en clave de competencia conlleva la incorporación de actividades formativas, que planteen demandas más allá de la reproducción de contenidos teóricos. Es así como se van incorporando tareas de diverso tipo y metodologías orientadas a la aplicación y transferencia (aprendizaje basado en problemas, proyectos, casos, etc.). Como será desarrollado a lo largo de los próximos capítulos, la presente investigación concibe la evaluación como un recurso didáctico con capacidad de promover aprendizaje. En consecuencia, ubicamos en este mismo bloque todo el trabajo de revisión e innovación que sobre las prácticas evaluativas está teniendo lugar².
- ✓ *Recursos didácticos.* La generalización de las tecnologías de la información y la comunicación en la sociedad dota de una amplia variedad de recursos la enseñanza y el aprendizaje. Su aplicación ofrece nuevas oportunidades y fomenta la innovación a niveles complementarios (planificación, realización de las tareas, gestión y seguimiento, evaluación y feedback, interacción y comunicación, etc.).

● ***Cambios en la delimitación del ámbito de responsabilidad formativa.***

De acuerdo con Benito y Cruz (2005) la “Educación Superior Europea supone asumir la existencia de unos objetivos de aprendizaje, o resultados, comunes a cada uno de los niveles educativos establecidos (grado y postgrado), que deberán atender a las necesidades planteadas por la sociedad actual” (p. 13). Internamente también, la definición de perfiles profesionales de egreso dentro de cada titulación e institución formativa hace más necesario que nunca el trabajo colegiado y la coordinación docente (Bonsón, 2009; Caurcel & Morales, 2008; M. Martínez & Viader, 2008). Definido en términos de competencia, el aprendizaje previsto integra teoría y práctica desde una perspectiva interdisciplinar y globalizadora, de modo que precisa de procesos de coordinación vertical y horizontal que favorezcan su consecución graduada y progresiva.

² Las implicaciones evaluativas sujetas al proceso de convergencia serán examinadas en el capítulo cuarto, al tratar los referentes para la definición de una teoría de la evaluación en el contexto universitario actual.

Desde el punto de vista docente, esta reconfiguración de las finalidades formativas modifica su ámbito de incumbencia y le lleva a exceder los límites de su propia disciplina. Más que un área o materia específica, se entiende que las contribuciones deben orientarse al perfil profesional global confeccionado y verificado en una determinada institución (L. C. Contreras, 2007; Sayós, 2010). Así lo manifiestan Bolarín, Moreno y Porto (2013):

Se precisa de formas de colaboración, cooperación y coordinación entre docentes, asumiendo la importancia y la pertinencia de relacionarse profesionalmente con el objetivo de plantear actuaciones conjuntas y compromisos de cara a asumir los nuevos retos educativos, orientados a la enseñanza y aprendizaje de competencias. Asumiendo, además, que el hecho de compartir estas experiencias con nuestros compañeros docentes constituye una forma válida y efectiva de desarrollo profesional y organizativo. (p. 446).

En definitiva, los cambios proyectados sobre el rol docente configuran una nueva imagen de la enseñanza más amplia y compleja. El replanteamiento de las competencias a desarrollar por el alumnado en su trayectoria formativa ha provocado la revisión de las propias competencias docentes del profesorado (Morales, 2008c), cuestión sobre la que profundizaremos en el segundo apartado. Para Bozu y Canto (2009),

estamos pues ante un supuesto perfil europeísta del profesorado, un perfil basado en competencias, con una nueva práctica docente, este profesional requerido sería un buen conocedor de las metodologías y didácticas activas (aprendizaje basado en problemas, aprender a aprender, dinámicas de grupo, acción tutorial, liderazgo etc.). Mucho tendrán que cambiar los docentes y nuestras universidades, para lograr tales objetivos, grandes dosis de innovación, ilusión y formación para llevar a cabo tan ardua tarea. (p. 90).

En esta imagen aún deben incorporarse las exigencias derivadas de la gestión, la investigación y la transferencia, previamente definidas. Su consideración conjunta, ha llevado a Galán y Rubalcaba (2007) y a V. Álvarez et al. (2009) a delimitar

distintos perfiles entre el profesorado universitario, según su orientación preferente a servicios comunitarios, docentes, institucionales o de investigación.

Como se expone en el próximo apartado, la respuesta solvente y equilibrada a este cúmulo de exigencias tiene lugar en un contexto académico-institucional y social que no siempre facilita su consecución o que incluso la obstaculiza.

1.2 Principales dificultades en el desempeño del cometido que se atribuye al profesorado universitario en el panorama actual.

La docencia representa una profesión compleja, desempeñada en un contexto y unas condiciones que también lo son. Según la nueva definición adoptada, constituye uno de los medios articulados desde la universidad para contribuir a la construcción del aprendizaje discente (Morales, 2006a) y puede ejercer un papel positivo o negativo sobre el mismo (Zabalza, 2006). Del mismo modo, la contribución docente es permeable a los factores académico-institucionales y sociales que la rodean, con capacidad de apoyarla u obstaculizarla según se produzcan.

La revisión de la literatura específica permite identificar con amplio margen de acuerdo una serie de factores o circunstancias que clásicamente dificultan el desempeño de los roles atribuidos al profesorado. A los propósitos del presente trabajo resultan de particular interés aquellas que afectan a la función docente (ver tabla 3), agudizadas en los últimos años por influencia indirecta de los cambios pedagógicos y estructurales vinculados al proceso de convergencia:

Tabla 3

Factores sociales y académico-institucionales que dificultan el ejercicio docente

PRINCIPALES DIFICULTADES LIGADAS A LA DOCENCIA UNIVERSITARIA

- Complejidad e intensificación de la enseñanza universitaria.
 - Indefinición profesional de la enseñanza universitaria y voluntarismo.
 - Dificil conciliación de los roles docente e investigador.
 - Devaluación de la docencia.
 - Privacidad de la enseñanza, individualismo y fragmentación del currículo.
 - Falta de medios y recursos y condiciones laborales poco favorables.
-

Este cúmulo de factores afecta en el ejercicio de la enseñanza universitaria y representa parte del contexto específico de la profesión. En consecuencia, es preciso valorar brevemente su influencia:

- El recorrido efectuado por el anterior apartado define para el profesorado universitario una gama de incumbencias que recorre cuatro espacios de actividad diferenciados. Villa (2008), Knight (2005) o Zabalza (2002) hablan de **intensificación** para referirse al incremento de tareas y funciones, en especial, desde que en la década pasada se emprendiera el reto de la consolidación del EEES. Fonseca (2010) lo expresa en los siguientes términos:

La responsabilidad social de las universidades se ha incrementado y diversificado. La función docente, siendo de importancia capital, se encuentra en desventaja. El profesorado tiene el compromiso de atender al constante perfeccionamiento en su propia área del saber, a las innovaciones tecnológicas y metodológicas y a las necesidades de formación que éstas plantean, a producir investigaciones de alto impacto, a la transferencia y difusión del conocimiento en sus distintas vertientes que incluyen ahora también la misión emprendedora, a la movilidad para conocer distintos entornos internacionales, a una gestión universitaria que permita todo lo demás. (p. 46).

Esta sobrecarga a menudo se acompaña de sentimientos de frustración ante la falta de reconocimiento de la labor docente por parte de alumnado y/o representantes institucionales (Monereo & Domínguez, 2014).

- Una de las dificultades con mayor reflejo en la investigación específica y la literatura es la **indefinición profesional de la docencia universitaria** (Cabrera & La Nasa, 2005; Feixas, 2002; A. Fernández, 2008; Kember, 1997; Monereo, 2014a), también denominada desprofesionalización docente (Zeichner, 2010). Su incidencia está motivada por las condiciones de acceso a la profesión, que se centran en la formación disciplinar, sin requerir una capacitación didáctico-pedagógica específica (J. M. Esteve, 2006; Mas, 2011, 2012; Montero, 2006). Como plantea Valcárcel (2003), la formación previa exige la superación de estudios de grado y postgrado en el ámbito de la titulación en la que se vaya a ejercer, y los programas de formación

inicial o introducción a la docencia no son obligatorios, a menos que la institución contratante así lo imponga. En consecuencia, el profesorado tiende a identificarse como experto en la materia, más orientado a la transmisión y la difusión del conocimiento que a la misión didáctica y pedagógica (Gros & Romaná, 2004; Parcerisa, 2010).

Los cambios acontecidos en los últimos años sobre el rol docente no han hecho sino reforzar el efecto de esta deficiencia, intensificando la demanda de formación específica para el ejercicio profesional de la enseñanza (V. Álvarez et al., 2011; Caballero, 2013; Delbanco, 2012; Mateo, 2012; Monereo, 2013, 2014a; Torra et. al., 2013). Las respuestas a esta circunstancia se van generando lentamente pero, entre tanto, la actividad docente se impregna de efectos adversos:

- ✓ Artificial simplificación de la función docente, erigida desde la presunción de que el dominio disciplinar basta para ejercerla. Sin embargo, enseñar es una tarea compleja que además exige comprender “cómo aprenden los estudiantes cuyo aprendizaje se ha de guiar, manejar bien los recursos de enseñanza que se adecuen mejor a las condiciones en que ha de realizar su trabajo, etc.” (Zabalza, 2002, p. 111).
 - ✓ La ausencia de requisitos de capacitación docente en el acceso a la enseñanza impregna la formación de un halo de voluntarismo, poco compatible con la identidad profesional del profesorado universitario (Michavila, 2005; Santos, 2010a; Torra et al., 2013). En palabras de M. M. Martínez, García y Quintanal (2006, p. 186), “la formación psicopedagógica del profesorado universitario ha sido tradicionalmente el producto de una voluntaria autoformación y de una socialización no consciente”, basada principalmente en la imitación y evitación de referentes considerados adecuados e inadecuados.
 - ✓ Reproducción de patrones docentes tradicionales, por ausencia de una fase específica de capacitación en la que se construyan referentes alternativos (Monereo, 2013; Tejedor, 1999). Su conocimiento sobre aprendizaje y enseñanza es principalmente vivencial, construido “a partir de la experiencia docente y de los modelos de socialización secundaria que les hacen repetir, sobre todo en el comienzo de sus carreras docentes, los esquemas de enseñanza de sus antiguos profesores” (Serrano, 2010, p. 271).
- La **conciliación de docencia e investigación** representa una dificultad clásica reactivada por los cambios estructurales sujetos a la convergencia, en especial, por

la generalización de los procesos de acreditación. Como resumen Tomás et al. (2012), la tensión resultante se manifiesta a dos niveles: en la convivencia o el grado de integración de ambas funciones en la vida académica y en el reconocimiento desigual que obtienen en los procesos de evaluación.

La confluencia de docencia e investigación representa un requisito ineludible para el profesorado universitario (A. Brew, 2003, 2007; Kindelán, 2013; Skelton, 2007). Sin embargo, su conjunción puede adoptar formas diferentes según se interprete una relación de superioridad y subordinación, enriquecimiento mutuo o plena integración entre ambas (Roberston & Bond, 2008). Los aspectos prácticos y funcionales de su unión, también suscitan dificultades evidentes. Vidal y Quintanilla (2000), por ejemplo, concluyeron en su estudio que el profesorado manifiesta dificultades para investigar cuando asume una carga elevada de docencia; en otros casos el desajuste apunta en sentido inverso, siendo la investigación la que, presuntamente, resta atención a la labor docente (Gibbs, 2008; Monereo & Domínguez, 2014; Santos, 1999).

Las conclusiones que arroja la literatura sobre esta convivencia revelan posturas muy diferentes al respecto (Villa, 2008). Según Hernández (2002) las posiciones principales pueden resumirse a tres: la de quienes confían en una correlación positiva de ambas actividades, la de quienes sostienen que dicha correlación resulta negativa y la de quienes reivindican la independencia de ambas tareas y su desarrollo por parte de agentes especializados. Ante tal divergencia de posiciones, la investigación sobre la propia docencia representa una alternativa emergente que, además de la conciliación, favorece la transferencia (A. Brew, 2007; Chocarro et al., 2013).

El reconocimiento desigual de ambos roles en los procesos de acreditación, también representa una fuente de controversia (M. Martínez, 2006; Parcerisa, 2010). En el contexto español, el acceso a las figuras de Profesor Contratado Doctor, Profesor Ayudante Doctor o Profesor de Universidad Privada está supeditado a una evaluación de méritos profesionales, efectuada de acuerdo al Programa de Evaluación del Profesorado para la contratación (PEP) (Agencia Nacional Evaluación de la Calidad y Acreditación-ANECA, 2007). Acaso por prestarse a una cuantificación más sencilla y directa, en dicha evaluación la investigación está sujeta a una mayor compensación que los méritos docentes o de gestión. El resultado es una orientación estratégica a la investigación por parte del profesorado (Tomás et al., 2012; Villa, 2008) y la devaluación de la docencia. Sobre ella repara la próxima dificultad.

- La indefinición profesional que acompaña a la docencia, y su desventaja acreditativa con respecto a los méritos de investigación han provocado la **devaluación de la docencia** (Fairweather, 2005; Knight, 2005). Esta devaluación aparece descrita desde dos perspectivas en la literatura. La primera de ellas tiene que ver con la proliferación de nuevas modalidades didácticas basadas en la tecnología de la información y la comunicación. Más rentables y económicos, según Paris (2013) los espacios de formación en línea restan mérito a la labor docente. La segunda está mucho más extendida y tiene que ver con el menor reconocimiento de la enseñanza frente a la investigación lo que en la práctica, imposibilita la realización de una carrera profesional docente (Gibbs, 2008; Inzunza, 2008; Ramsden, 2003). Así se constata también en los ránquines que clasifican la calidad de las instituciones universitarias; mientras que algunos de ellos se basan exclusivamente en la producción investigadora (Buela-Casal, Bermúdez, Sierra, Quevedo-Blasco, & Guillén-Riquelme, 2012), otros incorporan distintas dimensiones, pero preservando su supremacía (Times Higher Education, 2013³).
- Concebida como actividad solitaria y discrecional (Cid-Sabucedo et al., 2009), otra de las dificultades que acompaña tradicionalmente a la enseñanza es su **carácter privado** (Brockbank & McGill, 2002; Zabalza, 2002). Éste se traduce en un cierto individualismo que inhibe posibilidades de crecimiento y genera sensación de soledad (Bolarín et al. 2013; Monereo, 2013; Monereo & Domínguez, 2014). Desde la perspectiva curricular, contribuye a la descoordinación y a la fragmentación del currículum (Briggs, Stark, & Rowland-Poplowsky, 2003; Knight, 2005).
- La **falta de medios y recursos** representa una dificultad general en la educación superior y el profesorado es uno de los colectivos que más de cerca la vivencia. El incremento del número de estudiantes y de la diversidad de perfiles de acceso ha tenido lugar progresivamente, aunque no se ha acompañado de medios que posibiliten su atención oportuna (Apodaca & Grao, 1997; Santos, 1999). En este contexto se produce también el empeoramiento de las propias condiciones laborales del profesorado, en lo que Feixas (2002) caracteriza como el paso del elitismo a la precariedad laboral. “La situación de la facultad tradicional, con empleo a tiempo completo en un departamento y con una carrera bien definida que incluye la tenencia

³ En este prestigioso ranking mundial de universidades se preserva un 30% de la valoración global para la calidad de la docencia y otro 30% a la investigación (volumen, ingresos y reputación). El desnivel se produce cuando, en estrecha conexión con la labor investigadora, se imputa otro 30% para las citas o la influencia en la literatura y un 2,5% adicional para la innovación industrial.

y protección de la libertad académica, está desapareciendo” (Paris, 2013, p. 17). Esta inestabilidad y precariedad, ejerce una repercusión negativa sobre los procesos de colegialidad, formación y coordinación (Knight, 2005; Villa & García, 2014), reforzando el individualismo anteriormente señalado.

Los retos enumerados en el primer apartado así como la definición de los perfiles profesionales en la que se adentra el siguiente, están sujetos a estas circunstancias. La articulación de medidas para mejorarlas paulatinamente representa uno de los factores críticos que condiciona la consolidación definitiva del EEES (J. M. Escudero, 2008; Michavila, 2005; Rué, 2007).

2. Perfiles profesionales del profesorado en el EEES

El cambio de escenario descrito en el primer apartado y su incidencia en el cometido del profesorado universitario ha dado lugar a la emergencia de numerosas propuestas de revisión y sistematización de su perfil profesional, elaboradas desde perspectivas diferentes. Por una parte, la literatura y la investigación específica han centrado su atención en la identificación de los rasgos y condiciones que deben acompañar al profesorado para responder a los retos de la convergencia. Por otra parte, ese mismo proceso de reflexión y debate ha sido replicado internamente en el seno de las instituciones formativas para responder a circunstancias prácticas inmediatas, como el establecimiento de criterios para la gestión de personal o la articulación de los mecanismos de formación y actualización necesarios.

Entendemos por perfil profesional el “conjunto de competencias que identifican la formación de una persona, para asumir en condiciones óptimas las responsabilidades propias del desarrollo de funciones y tareas de una determinada profesión” (Bozu & Canto, 2009, pp. 89-90). Pavié (2011) lo describe como el “conjunto de los atributos de un rol o función profesional en virtud de competencias asociadas a las prácticas de la profesión” (p. 76). Representa una determinada imagen del tipo de profesional al que se aspira y en ella se entrelazan rasgos personales, sociales y profesionales (Feixas, 2008; Isaeva, 2007; Tejada, 2009).

Siguiendo la filosofía estructural y pedagógica del EEES, estos perfiles se definen adoptando las competencias como unidad básica de referencia. La ambigüedad y la

divergencia de interpretaciones que suscita dicho término, en función de los referentes culturales y su ámbito de aplicación, ha sido ampliamente constatada (Bonsón, 2009; A. Díaz, 2006; M. J. García, 2010; Mulder, Gulikers, Biemans, & Wesselink, 2009; Tobón, 2010; Viso, 2010) y no es objeto de este trabajo profundizar en ella. En consecuencia, desde una perspectiva práctica adoptamos como referencia la definición empleada en el Proyecto Tunning (J. González & Wagenaar, 2003), por ser probablemente la más influyente de las manejadas en el ámbito específico de la formación universitaria y por aproximarse a otras propuestas ampliamente referidas fuera del mismo (por ejemplo, las de Le Boterf, 2001; Perrenoud, 2004).

En el Proyecto Tunning, las competencias representan una combinación dinámica de atributos- con respecto al conocimiento y su aplicación, a las actitudes y a las responsabilidades- que describen los resultados del aprendizaje de un determinado programa, o cómo los estudiantes serán capaces de desenvolverse al finalizar el proceso educativo. (J. González & Wagenaar, 2003, p. 280).

Aunque esté orientada a las competencias precisadas en los planes de estudio para las/os estudiantes, esta conceptualización ofrece un esquema transferible a las que integran los perfiles profesionales del profesorado. Así lo entiende Saravia (2011), que define las competencias como “unidades de aprendizaje y demostración de desempeño exitoso tanto en los estudiantes como en el profesorado” (p. 10). En consecuencia, caracterizamos las competencias docentes como la respuesta eficaz y responsable a los cometidos atribuidos al profesorado en el complejo y cambiante contexto universitario, por medio de la integración de recursos de naturaleza diversa. Como en los resultados de aprendizaje discentes, en ellas también cabe la distinción de dos clases: las competencias básicas, genéricas o transversales (aquellas vinculadas a ámbitos profesionales diferentes) y las específicas (propias de una profesión en particular, en este caso, la enseñanza universitaria) (J. González & Wagenaar, 2003; Villa & Poblete, 2007).

2.1 Caracterización general de los perfiles profesionales propuestos para el profesorado en la era de la convergencia.

Las distintas propuestas de perfil profesional descritas para el profesorado en los últimos años muestran entre sí rasgos comunes y diferenciales. Elaboradas desde una perspectiva teórica general o en respuesta a las necesidades de una determinada institución formativa, se aprecian entre ellas distintos planteamientos y vías de aproximación para alcanzar un mismo objetivo: definir cuáles deben ser las competencias del profesorado en el actual marco formativo.

Para tomar conciencia de la dirección general que han adoptado los intentos de sistematización efectuados, hemos llevado a cabo una recopilación de algunas de las propuestas más significativas publicadas en el entorno próximo de la presente investigación. Sin ánimo de exhaustividad, la muestra conformada integra un total de 18 perfiles, seleccionados de acuerdo a cuatro criterios elementales:

- *Perfiles profesionales basados en la delimitación de competencias.* En línea con el planteamiento descrito en el apartado anterior, la búsqueda se orienta a perfiles profesionales definidos en términos de competencia y, excepcionalmente, estándares. Este primer criterio deja fuera de la selección aquellas propuestas centradas en la descripción de dimensiones, cualidades, rasgos o componentes aislados de la enseñanza, en los que no necesariamente se produce la integración de recursos que exigen las competencias (por ejemplo, las ideadas por Beltrán & Pérez, 2005; De Juanas, 2011; M. M. Martínez et al., 2006).
- *Perfiles profesionales específicamente orientados al profesorado universitario.* El trabajo con alumnado adulto, la finalidad formativa de capacitación profesional o el compromiso con la investigación y la transferencia son algunas de las particularidades de la enseñanza universitaria, frente a la que tiene lugar en otras etapas educativas (Feixas, 2002; Saravia, 2011; Zabalza, 2006). Para Davenport (2006) el profesorado universitario forma parte del sector de las/os trabajadoras/es del conocimiento, caracterizado por desempeñar una tarea desestructurada de difícil sistematización en la que se produce una mayor apreciación de la autonomía. Esta decisión ha dejado fuera de la selección otras propuestas interesantes, bien por centrarse en etapas educativas preuniversitarias, o por orientarse a la enseñanza en general, sin aterrizar en ninguna etapa educativa en particular (entre ellas se encuentran las de Cano, 2007; J. M. Escudero, 2006; Gairín, 2011; Perrenoud, 2004; Scriven, 1998; cit. en Cano, 2007; Viso, 2010).

- *Perfiles profesionales para el profesorado universitario en entornos presenciales.*
Cada vez más extendidos y generalizados, los contextos formativos virtuales suman a las particularidades anteriores otras propias, relacionadas con el medio y las peculiares formas de interacción que en él se producen. En consecuencia, no han sido considerados en el análisis los perfiles elaborados para el profesorado involucrado en entornos formativos no presenciales o mixtos (por ejemplo, los de Campos, Ligia, & Solano, 2010; Da Silva, 2007; Gallego, 2007; Unigarro & Rondón, 2005).
- *Perfiles profesionales específicamente orientados al contexto formativo del EEES.*
Para configurar una muestra de perfiles mínimamente homogénea y próxima al contexto de la investigación, se ha limitado la selección de propuestas al marco geográfico y cronológico de la convergencia. Los perfiles analizados han sido publicados entre 2003 y 2014 y en referencia al contexto universitario de los estados miembro, preferentemente España. Este criterio deja fuera de la muestra contribuciones importantes efectuadas fuera del continente europeo, por ejemplo, en el ámbito formativo mexicano (Ayala, 2008; cit. en Tejada, 2009; B. García, Loredó, Luna, & Rueda, 2008; Rueda, 2009), australiano (AITSL- Australian Institute for Teaching and School Leadership, 2012) o colombiano (Camargo & Pardo, 2008).

Resumidos en el **ANEXO 01**, el análisis de los perfiles recopilados ofrece una visión general de las competencias requeridas al profesorado universitario desde el comienzo del nuevo siglo. No se trata de una síntesis pormenorizada, pero el contraste y la consideración conjunta de las propuestas arroja ciertas conclusiones. Las más relevantes para comprender la dirección general que ha tomado la delimitación de perfiles profesionales para el profesorado universitario son las señaladas a continuación:

- La revisión de los perfiles propuestos en la muestra confirma la preeminencia del rol docente frente a la gestión, la investigación y la transferencia. En la mayor parte de las propuestas las competencias formuladas se centran principalmente en el cometido docente, aquel espacio de actividad más próximo al alumnado (por ejemplo, Alegre & Villar, 2006; A. Fernández, 2008; Torra et al., 2013; Zabalza, 2003). Este protagonismo también se advierte en los perfiles integrados, en los que las competencias incluidas refieren a distintos espacios de actividad del profesorado sin limitarla a la enseñanza (V. Álvarez et al., 2011; Caurcel & Morales, 2008; Ortega, 2010). No son frecuentes las propuestas como las de Mas y Ruiz (2007) o

Sagasta et al. (2006⁴), que delimitan competencias diferentes para las distintas funciones a desempeñar por el profesorado.

- Las propuestas analizadas dan cuenta de perfiles transferenciales, flexibles y polivalentes, en línea con las recomendaciones de Valcárcel (2003) y Bozu y Canto (2009). No representan un listado cerrado de logros predefinidos y estables que pueda entrenarse y acreditarse, garantizando la efectividad del personal que los desarrolla (B. García et al., 2008). Ofrecen más bien un mapa general de los principales cometidos que deberá solventar el profesorado en su complejo y cambiante contexto profesional, por medio de la movilización armónica y creativa de recursos de naturaleza diversa. Así lo requiere la respuesta efectiva a los retos formativos del EEES y la elección de las competencias como eje central para la delimitación del rol docente y discente. Pavié (2011) lo resume de este modo:

Un profesional competente, e incluimos aquí al profesor, debe ser capaz de transferir y adaptar, en el marco de su desempeño laboral, uno o varios esquemas de actividad a diversas situaciones o problemas que se le presentan, ya que la competencia no es tanto una característica del trabajo en sí, sino de quienes lo ejecutan bien. (p. 78).

- Los perfiles seleccionados revelan la divergencia de interpretaciones que produce en la práctica el término competencia. Estas discrepancias se evidencian, al menos, a dos niveles:
 - ✓ *En el enunciado de las competencias.* Las recomendaciones para la formulación de competencias sugieren que se conjugue un verbo de desempeño, un objeto, una finalidad y la condición o condiciones de referencia aplicables a cada caso (Tobón, 2010). Lejos de este esquema, en los perfiles analizados las competencias aparecen caracterizadas como acción (redactadas con verbo), como producto o resultado (encabezadas por un sustantivo) o como rol o cualidad docente (resumidas mediante adjetivos o calificativos) (ver **ANEXO 01**). En cualquiera de las alternativas, esta delimitación inicial se amplía y complementa con una descripción más o menos detallada de su sentido y significado, aunque con grandes diferencias

⁴ Esta propuesta propone competencias diferenciadas para las funciones docente, gestora e investigadora. No ha sido incluida en la muestra por la imposibilidad de acceder a la definición completa del perfil.

en su nivel de desarrollo: pequeñas aclaraciones que completan el enunciado (Caurcel & Morales, 2008; Ortega, 2010), descripciones elaboradas que contribuyen a una mayor profundización (Alegre & Villar, 2006; Zabalza, 2003) o propuesta de indicadores que precisan el alcance de la competencia a la que representan (V. Álvarez et al., 2011; I. González & López, 2010; Mas & Ruiz, 2007; Saravia, 2011; Tigeelar, Dolmans, Wolfhagen, & Van der Vleuten, 2004; Torra et al., 2012).

- ✓ *En el sentido y significado atribuido a las mismas.* En la definición asumida las competencias exigen la combinación de saberes conceptuales, procedimentales y actitudinales. Sin embargo, en las propuestas de Valcárcel (2003) o Tigeelar et al. (2004) se asimila la competencia con el dominio teórico o disciplinar, presentándolo como un resultado en sí mismo. S. Fernández (2003) también integra el conocimiento disciplinar en su propuesta, pero amplía levemente la perspectiva al hacer referencia a su gestión. Este planteamiento refleja una artificial limitación del significado de las competencias que no se produce en propuestas más recientes, donde las acciones, roles o resultados formulados exceden el mero conocimiento (necesario, pero insuficiente por sí mismo).
- La revisión efectuada revela un claro predominio de las competencias específicas frente a las genéricas en los perfiles del profesorado. Pavié (2011) considera que “quizá sea un poco pretencioso indicar que existen competencias exclusivamente docentes ya que en sus enunciados tales competencias no se diferencian mucho de otras que tienen un campo de aplicación distinto [...]” (p. 77). Sin embargo, las propuestas y su descripción se sitúan muy próximas al marco específico de actuación docente, hasta el punto que resulta complicado identificar las competencias genéricas (ver **ANEXO 01**). Para hacerlo, no basta con atender al enunciado de la competencia, sino que debe considerarse también el sentido que se le confiere en su descripción. Así por ejemplo, competencias aparentemente transversales, como la comunicativa o la relativa al manejo de las TIC, aparecen reinterpretadas como recursos pedagógicos para la interacción didáctica en la mayor parte de los perfiles.
- La revisión vertical de los perfiles seleccionados ofrece reflejos evidentes del paradigma formativo del EEES. El más claro de todos ellos es la insistencia en el aprendizaje de las/os estudiantes, denominador común que subyace a todas las propuestas de forma más o menos explícita. Atribuciones clásicamente docentes, como la planificación o la articulación didáctica de métodos y recursos, incluyen

referencias directas al papel del alumnado y a las repercusiones en su actividad y aprendizaje.

- Sin perjuicio de la anterior conclusión, no es menos cierto que, en los perfiles considerados, siguen presentes las principales dimensiones señaladas para la competencia docente mucho antes de que se iniciase el proceso de convergencia. Olivia y Henson (1980) proponían hace más de tres décadas los conocimientos básicos y las destrezas comunicativas, técnicas, administrativas e interpersonales como principales ámbitos de competencia docente, sistematización que sigue siendo aplicable en la actualidad. En propuestas elaboradas fuera del ámbito de influencia del EEES también se acusa el protagonismo del aprendizaje y los rasgos generales enumerados hasta el momento (predominio del rol docente y de las competencias específicas, discrepancias en la formulación y concepción de las competencias y flexibilidad y polivalencia de las propuestas).

El resumen efectuado deja tras de sí una última conclusión evidente. No existe un único perfil para el profesorado, incluso cuando la búsqueda se acote al ámbito específico de la formación universitaria (V. Álvarez et al., 2009; Feixas, 2002; M. M. Martínez et al., 2006; Monereo & Badía, 2011). La cantidad y variedad de propuestas constatada refleja aproximaciones parciales a una realidad compleja, susceptible de ser interpretada desde prismas diferentes. Como ya se advertía en apartados anteriores, el rol docente definido en el EEES representa el denominador común que aproxima las propuestas, pero no las uniformiza ni lo pretende⁵. En el contexto formativo europeo, existe un enfoque u orientación general para el ejercicio de la enseñanza universitaria, no un único modelo o perfil profesional.

2.2 Las competencias profesionales del profesorado universitario: perspectiva institucional de la Universidad de Deusto.

Admitido que en el paradigma formativo de la convergencia caben distintos perfiles de la profesión docente, este nuevo apartado detalla aquel que se asume como referencia en el presente trabajo.

⁵ También en este aspecto se hace evidente la filosofía de la sintonización o armonización desde la que se articulaba el Proyecto Tunning (J. González & Wagenaar, 2003).

Igual que se definen las competencias profesionales que debe desarrollar el alumnado a lo largo de su formación básica, es deseable que cada institución y/o titulación se detenga en la delimitación del perfil del profesorado que mejor pueda contribuir a ese logro y al resto de finalidades estratégicas previstas. Nuestra investigación se encuentra enraizada en el contexto institucional de la Universidad de Deusto, donde también efectúa su aproximación empírica. En consecuencia, adopta como referencia el perfil de profesorado definido en dicha institución para encauzar los procesos de evaluación y acreditación de la calidad de la docencia.

Fruto de un proceso de investigación y reflexión interna, Villa y García (2008, 2014) plantean un modelo de estándares de calidad y excelencia docente para su implementación y desarrollo en los nuevos títulos de grado. La propuesta se centra específicamente en el rol docente del profesorado, aquel espacio de actividad en el que profundiza el presente trabajo. Consiste en un sistema de 17 estándares, ordenados en cinco subprocesos clave involucrados en la enseñanza (ver tabla 4). Esos estándares tratan de evaluar las competencias genéricas y específicas que debe poner en práctica el profesorado para desarrollar el proceso de enseñanza-aprendizaje eficazmente.

Tabla 4

Modelo de competencias docentes del profesorado de la Universidad de Deusto (Villa & García, 2008; cit. en Villa, 2008, pp. 201-203)

PROCESO DE DOCENCIA: DELIMITACIÓN DE SUBPROCESOS Y COMPETENCIAS	
SUBPROCESO DE PLANIFICACIÓN	
<i>Planificación.</i>	Definir y organizar los objetivos, contenidos, metodologías, recursos, materiales, sistemas de tutoría y evaluación que aseguren el aprendizaje autónomo y significativo por parte de las/os estudiantes.
<i>Manejo de las TICs.</i>	Planificar y organizar la incorporación de las TICs como medio de apoyo y soporte a los procesos de enseñanza y aprendizaje.
SUBPROCESO DE GESTIÓN PEDAGÓGICA	
<i>Gestión del tiempo.</i>	Capacidad de organizar de forma equilibrada la carga de trabajo del estudiante y del profesor, en coherencia con las prioridades y estimaciones planificadas y atendiendo a las necesidades del proceso de aprendizaje.
<i>Liderazgo.</i>	Capacidad de influir sobre individuos o grupos anticipándose al futuro y contribuyendo a su desarrollo personal y profesional.
<i>Relaciones interpersonales.</i>	Hace referencia a la relación positiva con las/os estudiantes, por medio de la escucha empática y de la expresión clara y asertiva, ya sea por vías verbales como no verbales.

PROCESO DE DOCENCIA: DELIMITACIÓN DE SUBPROCESOS Y COMPETENCIAS	
<i>Gestión del aprendizaje.</i>	Se centra en el desempeño de la tarea docente a través de la promoción, orientación y seguimiento del aprendizaje, por medio de recursos didácticos y personales. Todo ello, para favorecer el aprendizaje autónomo y significativo de las/os estudiantes.
<i>Creación de un clima favorable aprendizaje.</i>	Se trata de asumir un compromiso de exigencia y profundidad del aprendizaje y de calidad del trabajo del estudiante, para lo que se promueve un clima favorable a ello.
<i>Coaching / apoyo al estudiante.</i>	Comportamiento docente de apoyo al desarrollo académico y personal del estudiante, que se traduce en la ayuda y el asesoramiento para dar con respuestas ante situaciones de dificultad o bloqueo.
SUBPROCESO DE EVALUACIÓN	
<i>Evaluación del aprendizaje.</i>	Se define como la comprobación y comunicación del grado de consecución de los objetivos de aprendizaje establecidos en la asignatura, a lo largo del proceso y como resultado final del mismo.
SUBPROCESO DE REVISIÓN Y MEJORA	
<i>Autoevaluación.</i>	Comprobación y comunicación del grado de consecución de los objetivos de aprendizaje establecidos en la asignatura, a lo largo del proceso y como resultado final del mismo.
<i>Orientación a la calidad.</i>	Se entiende como el comportamiento para realizar y mantener un trabajo de calidad de acuerdo con las normas y gestionar por procesos utilizando indicadores de calidad para su mejora continua.
SUBPROCESO DE COLEGIALIDAD	
<i>Orientación a resultados.</i>	Comportamiento encaminado al logro de nuevos resultados con éxito en el ámbito académico, incrementando las tasas de permanencia, éxito, rendimiento e inserción laboral.
<i>Innovación pedagógica.</i>	Capacidad para crear o modificar al que supone una novedad y un valor añadido, dando respuesta a un problema o desarrollo.
<i>Compromiso institucional.</i>	Identificación con la misión universitaria y la responsabilidad social que ello conlleva.
<i>Compromiso ético.</i>	Hace referencia al comportamiento coherente en relación con los valores éticos. Se basa en el respeto a los derechos y obligaciones propios y ajenos, e implica el compromiso con los valores definidos en el proyecto universitario.
<i>Orientación al desarrollo profesional.</i>	Capacidad de modificar de forma flexible y crítica los esquemas mentales previos, a partir de la incorporación continua de información nueva que se emplea para comprender y transformar la realidad, promocionando el desarrollo personal y profesional.
<i>Trabajo en equipo.</i>	Hace referencia a la capacidad de integrarse y colaborar activamente en la consecución de objetivos comunes.

La sistematización expuesta, ofrece una visión de la enseñanza universitaria completa, actualizada y, en general, próxima a la que se desprende de la muestra de

perfiles analizada. Así lo refleja la tabla 5, que representa de forma sintética el respaldo que encuentran los 17 estándares formulados en otras propuestas surgidas en el entorno próximo.

Tabla 5

Correspondencia de los estándares de calidad y excelencia docente de la UD y los perfiles profesionales analizados

ESTÁNDARES DE CALIDAD y EXCELENCIA DOCENTE (Villa & García, 2008)

OTROS PERFILES	Planificación	Manejo de las TIC	Gestión del tiempo	Liderazgo	Relaciones interpersonales	Gestión del aprendizaje	Creación de un clima favorable al aprendizaje	Coaching / apoyo al estudiante	Evaluación del aprendizaje	Autoevaluación	Orientación a la calidad	Orientación a los resultados	Innovación pedagógica	Compromiso institucional	Compromiso ético	Orientación al desarrollo profesional	Trabajo en equipo
Torra et al. (2012)																	
UKPSTF ^a (2012)																	
Villa, Carbone, & Troncoso (2012)																	
V. Álvarez et al. (2012)																	
A. Medina et al. (2011)																	
Saravia (2011)									X								
Ortega (2010)									X								
I. González & López (2010)																	
Molenaar et al. (2009)																	
Tejada (2009)																	
Caurcel & Morales (2008)																	
A. Fernández (2008)									X								
Mas & Ruiz (2007)																	
Alegre & Villar (2006)																	
Tigelaar et al. (2004)																	
S. Rodríguez (2003)									X								
Valcárcel (2003)																	
Zabalza (2003)																	
f	16	7	4	1	12	15	6	11	14	9	1	0	9	5	5	11	11
% Representación	88.8	38.8	22.2	5,5	66,6	83.3	33.3	61.1	77.7	50	5,5	0	50	27.7	27.7	61.1	61.1

Nota: Las cuatro tonalidades de sombreado aplicadas a las competencias reflejan su grado de representación en la muestra de perfiles analizada. Las dos últimas filas recogen respectivamente, la frecuencia o el número de perfiles que contemplan cada competencia, y el porcentaje que dicha cifra supone sobre el total de las propuestas. Las competencias con una representación por debajo 25% en la muestra (muy escasa) aparecen sombreadas en rojo; aquellas presentes en más del 25% y menos del 50% de los perfiles (representación

escasa) se escriben sobre fondo morado; las competencias con una representación entre el 50% y el 75% (buena representación) aparecen sombreadas en verde; y, por último, las competencias con mayor representación en la muestra (más del 75%) se señalan sobre fondo azul. Asimismo, se ha incluido una X en aquellas propuestas que no incluyen expresamente la competencia referida a la evaluación del aprendizaje, núcleo central del presente estudio.

^a UKPTSF. UK Professional Teaching Standards Framework.

El perfil docente adoptado no representa la única alternativa existente, ni tampoco la más completa; sin embargo, ofrece un marco de referencia claro y contextualizado para avanzar en nuestros propósitos, y así se justifica en el próximo apartado.

3. Evaluación del aprendizaje: competencia docente clave del profesorado universitario

El perfil profesional docente seleccionado ofrece un mapa general de competencias muy amplio, en correspondencia con la complejidad de la enseñanza universitaria. Para favorecer una mayor exhaustividad en sus aportaciones, la presente investigación centra su atención en una de ellas solamente: la ***evaluación del aprendizaje***.

Incluida en el subproceso tercero, la evaluación del aprendizaje representa una de las competencias específicas integradas en el modelo de Villa y García (2008, 2014), dada su especial vinculación con el ejercicio profesional de la docencia. Tal y como aparece definida, hace referencia a la evaluación del aprendizaje de las/os estudiantes, el más clásico y representativo de los objetos de análisis en el proceso formativo, aunque no el único. La conceptualización articulada en el próximo capítulo presenta la evaluación como un proceso clave que puede ser ejercido por distintos agentes (docentes, discentes, expertas/os o agentes externos, representantes de la administración, etc.) y sobre objetos diversos (calidad de la oferta formativa, satisfacción de los agentes implicados, adecuación de los recursos, coordinación, calidad docente, etc.). Sin embargo, de todas las manifestaciones posibles, la presente investigación reduce la competencia evaluativa del profesorado universitario al marco de las evaluaciones que tienen lugar por iniciativa docente y sobre los procesos y productos del aprendizaje del alumnado. Esta delimitación aterriza en el espacio de actividad docente expresamente, sin negar que la enseñanza universitaria implique también el desarrollo de funciones ligadas a la investigación, la gestión y la transferencia.

Desde esta perspectiva, entendemos la competencia evaluativa del profesorado universitario como la respuesta eficaz y responsable a los dilemas y demandas que plantea en la práctica la evaluación del aprendizaje de las/os estudiantes (figura 5). Consiste en la producción de evaluaciones de calidad⁶ en el contexto real en el que tiene lugar el proceso de enseñanza-aprendizaje.

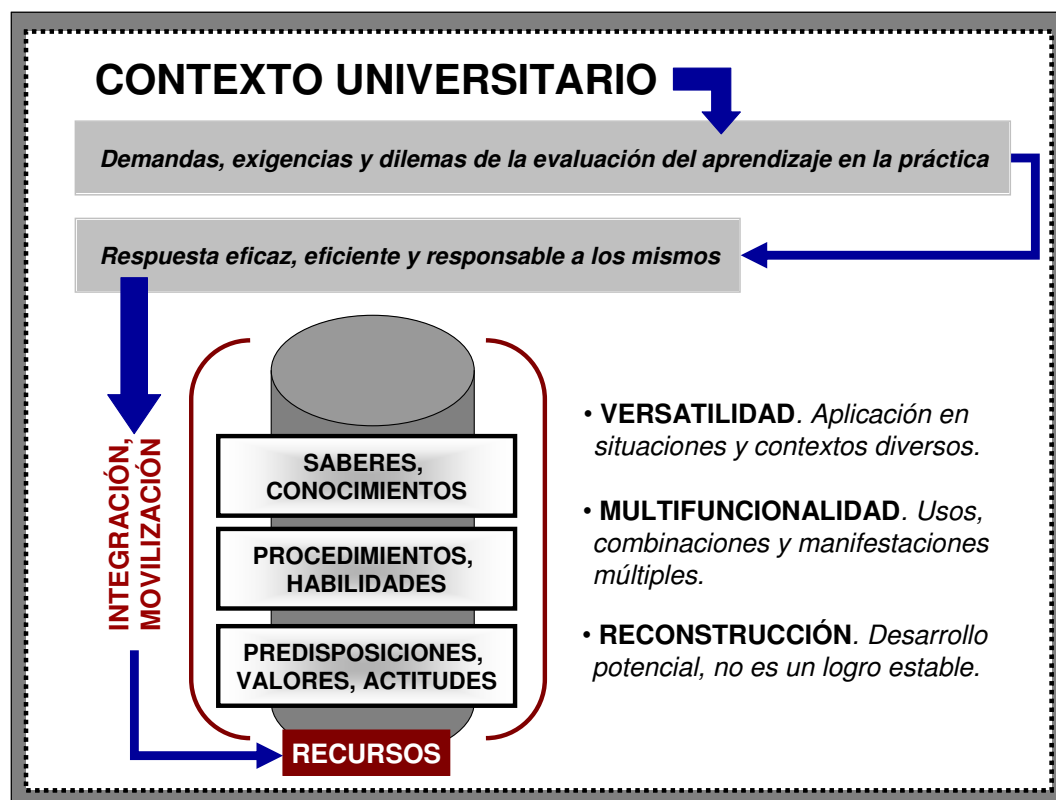


Figura 5. La competencia evaluativa del profesorado universitario: claves para su delimitación.

Toledo (2006) propone una conceptualización similar y la define en los siguientes términos:

La competencia evaluativa es la habilidad del docente para verificar el progreso de los alumnos, su capacidad para usar un amplio campo de conocimientos que le permite responder a las demandas. Es el buen uso del saber y del saber hacer en situaciones dadas. Es el saber intervenir, buscando los indicadores de que el sujeto

⁶ La evaluación de calidad es aquella que se ajuste a los parámetros de la teoría de la evaluación que se sostenga en cada caso. Ante la polisemia y diversidad natural de orientaciones que cabe esperar en este sentido, será objeto del capítulo cuarto la presentación y clarificación de lo que en esta investigación se considera una buena praxis evaluativa (la teoría evaluativa de referencia).

ha logrado múltiples objetivos educativos. También es su capacidad para concebir y generar situaciones evaluativas de acuerdo al nivel de los alumnos, con una visión amplia de los objetivos que estos deben lograr, señalando el progreso de cada uno en una aproximación formativa y de regulación de su aprendizaje. (p. 108).

Como cualquier otra competencia, la evaluación surge de la combinación armónica de recursos personales, sociales y profesionales de naturaleza diversa. Su activación tiene lugar en respuesta a situaciones y demandas cambiantes y complejas, en este caso, las suscitadas en la praxis evaluativa universitaria. Movilizados de forma flexible, los recursos que se conjugan en la misma responden a tres grandes categorías o dimensiones interrelacionadas (Interstate New Teachers Assessment and Support Consortium-INTASC, 2011; Trillo, 2005; UK Professional Teaching Standards Framework-UKPTSF, 2012; Villardón, 2006):

- La categoría teórica, que abarca conocimientos y saberes declarativos sobre el aprendizaje y su evaluación en sentido amplio (Toledo, 2005). De acuerdo con McMillan (2000) las múltiples elecciones que supone la delimitación y el desarrollo de dispositivos evaluativos, debe sustentarse necesariamente sobre esta base.
- La categoría procedimental, que se corresponde con lo que Biggs (2005) denomina los aprendizajes funcionales. Es la dimensión que conecta con la actuación y ejecución cualificada. Su objetivo es articular estrategias que trasladen a la práctica los planteamientos teóricos que se desprenden del saber declarativo en materia evaluativa (Trillo, 2005).
- La categoría actitudinal, conformada por las predisposiciones, principios y valores vinculados a la dimensión ética y profesional de la evaluación. Su concurrencia resulta especialmente relevante en la competencia que nos ocupa, ya que genera naturalmente sensación de poder (Batalloso, 2004) y es sensible a dilemas éticos y conflictos de interés en su puesta en práctica. Evaluar exige emitir un juicio sobre el nivel y la calidad del aprendizaje construido por las/os estudiantes; no se trata de una práctica neutral ni aséptica y plantea decisiones complejas y delicadas a quien la ejerce de forma responsable (Race, 2003; Santos, 2003; Stobart, 2010).

Pero como matiza De la Orden (2011a), la competencia no se identifica con la acumulación de saberes declarativos, procedimentales y actitudinales, sino con el

desempeño eficaz y eficiente de un papel o de una función en un ámbito o contexto determinado. Ello implica el dominio de las tres categorías referidas pero, ante todo, el uso experto e integrado de los recursos que las mismas proporcionan. Por su marcada naturaleza situacional, esta combinación de recursos da lugar a un repertorio amplio y variado de posibles usos y manifestaciones, caracterizado de forma general como multifuncionalidad (Cano, 2007; Unigarro & Rondón, 2005; Villa & Poblete, 2007; Viso, 2010).

En consecuencia, la competencia evaluativa no debe concebirse como un inventario de los saberes, procedimientos y componentes actitudinales necesarios para evaluar el aprendizaje discente en la universidad. Abierta y compleja, su aplicación puede tener lugar en situaciones y contextos diferentes y adoptar una gama de manifestaciones muy variada. En virtud del carácter reconstructivo reconocido a las competencias (Cano, Lleixà, Ion, & Tierno, 2010; Le Boterf, 2001; Pavié, 2011) puede desarrollarse a lo largo de la carrera docente en función de la cantidad y calidad de experiencias vivenciadas y de la reflexión que las mismas generen.

Sin negar la relevancia y pertinencia del resto de competencias que completan el perfil docente, en la presente investigación la evaluación del aprendizaje adopta una significación especial y llega a concebirse como competencia docente clave. Las evidencias que apoyan su protagonismo son abundantes:

- Representa un núcleo de interés clásico e intemporal por su potencialidad didáctica y el amplio espectro de repercusiones que acompañan a su ejercicio (académicas, sociales, personales o económicas) (Bonsón & Benito, 2005; Cid-Sabucedo et al., 2009; Race, 2003; Zabalza, 2003).
- Las referencias que avalan su relevancia se producen desde los niveles educativos básicos (J. M. Álvarez, 2000; Linn & Gronlund, 2000; De Ory & Ruiz, 2011; Martínez-Rizo & Mercado, 2015; J. Rodríguez, 2008; Villar, 2009). No representa una inquietud exclusiva del ámbito formativo universitario, sino que tiene impacto en todo el sistema educativo.
- Según el análisis efectuado, es una de las competencias docentes con mayor representación en la muestra de perfiles profesionales recabada para el profesorado universitario⁷ (ver tabla 5). Junto a la planificación y la gestión del aprendizaje, figura entre los componentes que más se repiten entre las propuestas.

⁷ Incluso en las cuatro propuestas en las que no se incorpora expresamente, la evaluación del aprendizaje puede suponerse presente en el desarrollo de otras competencias: en A. Fernández (2008) y Ortega

Este protagonismo hace que la evaluación del aprendizaje discente se señale como uno de los factores clave de la calidad institucional y una línea estratégica para su promoción y desarrollo (Gibbs, 2010; Kaneko, 2008; M. L. Pérez et al., 2000; Porto, 2010; Stiggins & Duke, 2008; Villa, 2004). En consecuencia, la mejora de los sistemas de evaluación debería constituir un elemento clave en los proyectos de promoción de la calidad docente al nivel institucional, no el producto de intervenciones puntuales y aisladas.

De forma más precisa, sostenemos que la investigación sobre la competencia evaluativa del profesorado universitario debe ser prioritaria y preferente por cinco motivos principalmente. Anticipados en la figura 6, su justificación tiene lugar en los próximos apartados.

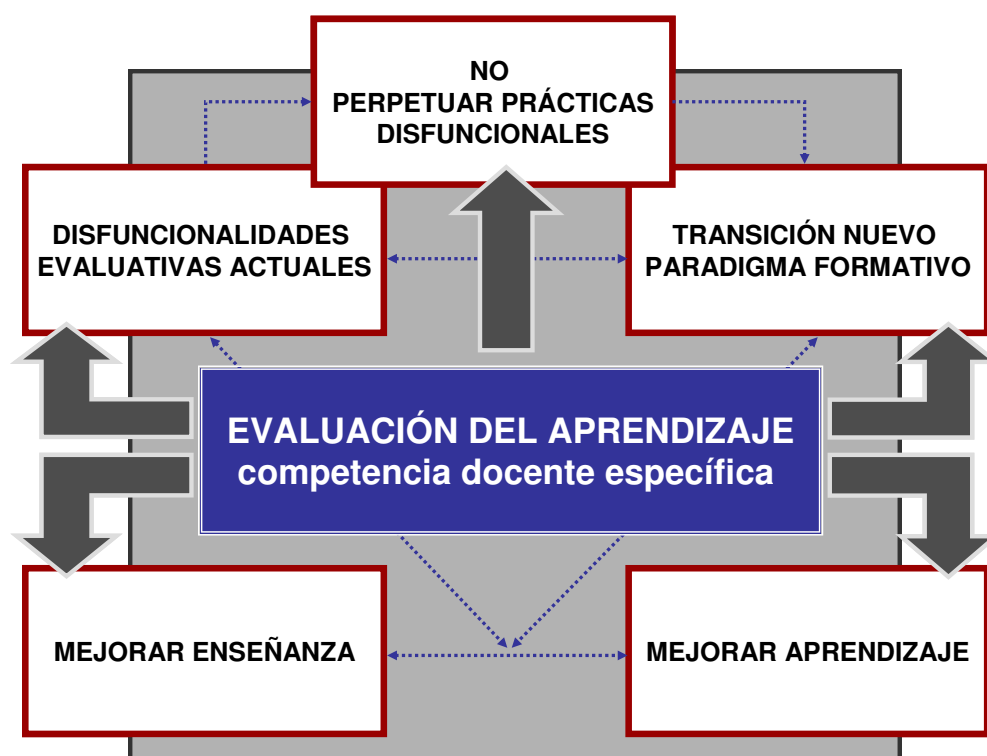


Figura 6. Razones para profundizar sobre la competencia evaluativa del profesorado.

(2010) en las relativas a la gestión del proceso de enseñanza y aprendizaje; en Saravia (2011), entremezclada con varias dimensiones de la competencia social; y en S. Rodríguez (2003) en el diseño, planificación y gestión curricular.

3.1 Las constatadas disfuncionalidades de los sistemas tradicionales de evaluación del aprendizaje en la universidad.

Como señala Race (2003), si las prácticas e instrumentos de evaluación funcionasen de forma correcta y satisfactoria, no habría necesidad de introducir innovaciones en las mismas. Sin embargo, no es esa la realidad que muestran los múltiples estudios efectuados sobre el terreno. Generalizada como práctica final y sumativa, “más que evaluar para aprender y comprender cómo y por qué aprenden lo que aprenden los niños, jóvenes y adultos en su proceso educativo, la evaluación de los aprendizajes parece haberse conformado con medir cuánto rinden o logran los estudiantes” (Román & Murillo, 2009, p. 5). Tal vez por ello, figura entre las competencias más controvertidas y peor valoradas en la actuación docente (Buckley, 2012; Flórez, 1999; Higher Education Founcil Council for England- HEFCE, 2014).

Desde esta perspectiva, la primera razón para centrar la atención en la competencia evaluativa del profesorado son las muestras evidentes de deficiencia y disfuncionalidad presentes en los sistemas de evaluación del aprendizaje universitario. El alumnado se beneficia de las evaluaciones bien planteadas y desarrolladas y resulta perjudicado por las evaluaciones de baja calidad (JCSEE, 2007). Así lo entiende también I. Álvarez (2008a) quien, incluso advirtiendo de los riesgos que entraña emprender innovaciones en materia evaluativa, insiste en la necesidad de hacerlo en vista de la situación actual.

El capítulo tercero del marco teórico se detiene en la enumeración de las principales deficiencias y dificultades evaluativas constatadas al describir el estado de la cuestión.

3.2 El valor estratégico de la evaluación en la mejora de la calidad del aprendizaje universitario.

Según se expone en el tercer capítulo, la evaluación del aprendizaje en el contexto universitario se ha visto reducida y limitada a una acción puntual y finalista que transcurre en paralelo al aprendizaje y la enseñanza. A efectos prácticos se corresponde con la calificación, aunque desde la perspectiva teórica se sostenga con amplio acuerdo que debe ser mucho más que eso. Así lo constatan Ibarra y Rodríguez (2010) efectuado un análisis sobre el discurso evaluativo dominante en diversas universidades españolas:

La primera conclusión, clara y evidente, que se desprende del análisis es que el discurso de la evaluación podríamos caracterizarlo como clásico o tradicional, es decir, en el sentido de equiparar evaluación a calificación/medición. A pesar de una clara necesidad percibida de reconocer otros propósitos (mejora, aprendizaje,...), éstos no se hacen evidentes. Consecuentemente, el aprendizaje, en la normativa reguladora de la evaluación en las universidades, no es reconocido como un eje central. (p. 400).

Esta perspectiva de la evaluación sólo responde a una de las finalidades que cabe atribuirle, reduciendo y limitando reseñablemente el proceso. De acuerdo con V. M. López (2004), “lo grave no es lo generalizada que está dicha confusión terminológica, sino el hecho de que las prácticas evaluativas (formativas, educativas), han desaparecido bajo la preeminencia y dominio de las prácticas calificativas” (p. 225).

Como hemos indicado en otros foros (Bilbao, 2011a, 2011b), lo habitual es concebir la evaluación como herramienta para medir el aprendizaje, no como un recurso para promocionarlo o estimularlo. “Tanto la evaluación *del* aprendizaje como la evaluación *para* el aprendizaje son esenciales, pero mientras que la primera está en vigor actualmente, la segunda no” (Stiggins, 2002, p. 758). El balance entre ambas formas de evaluación se encuentra muy desproporcionado a favor de la primera (Edwards, Turner, & Mokthari, 2008). Más que una cultura pedagógica de la evaluación, la reducción y limitación de sus funciones en la práctica se corresponde con una cultura del examen, centrada en la acreditación de productos finales (Dochy et al., 2002; Boud & Falchikov, 2007). En este contexto, emerge irremediablemente el protagonismo absoluto de las calificaciones (MacLellan, 2001; Race, 2003), en un fenómeno que comienza a manifestarse desde los niveles educativos básicos (M. A. Calatayud, 2004; D. Salinas, 2002; Stiggins, 2007).

Ante este cuadro de visiones y reacciones convencionales, emerge desde hace años y con amplio respaldo una nueva forma de entender el fenómeno evaluativo. Caracterizada por enfatizar su valor pedagógico y proponer una serie de principios generales para su explotación didáctica⁸, esta visión de la evaluación recibe denominaciones diferentes según los matices específicos a los que se aproxime. Frente a la tradicional evaluación *del* aprendizaje, se caracteriza de forma general como evaluación *para* el aprendizaje

⁸ Sobre ellos y su extrapolación al ámbito formativo universitario tendremos ocasión de profundizar en el capítulo cuarto.

(Assessment Reform Group-ARG, 1999; J. Chappuis, 2009; J. Chappuis, Stiggins, Chappuis, & Arter, 2012; S. Chappuis & Stiggins, 2002; Hargreaves, 2007; Harlen, 2006) u orientada al mismo (Carless, 2002, 2007; Carless, Joughin, Liu, & asociados, 2006; Carless, Joughin, & Mok, 2006; V. M. López & Palacios, 2012). Con independencia de la denominación precisa que adopte⁹, en la nueva perspectiva la evaluación comienza a considerarse como un recurso didáctico para la consecución de los objetivos de aprendizaje, y no como un mero instrumento para generar calificaciones.

Esta nueva visión reposa en la convicción de que la evaluación representa la más poderosa arma con la que cuenta el profesorado para incidir en las estrategias de aprendizaje de las/os estudiantes (Gibbs, 2003; Gibbs & Dunbar-Goddet, 2007). Las presiones asociadas a la evaluación y el protagonismo natural de las calificaciones, provoca que los dispositivos de evaluación empleados determinen de forma poderosa las estrategias de aprendizaje y el trabajo del alumnado (Brown, 2003b; Leal, 2004; Stobart, 2010; Vu & Dall’Alba, 2007; Wiliam, 2011). “Dado que el éxito se mide por las calificaciones recibidas, los estudiantes se esfuerzan por ser ganadores eficientes y eficaces de calificaciones” (Haigh, 2007, p. 458).

Por todo ello, la evaluación se constituye como herramienta esencial para estimular la construcción de aprendizajes de calidad (Bonsón & Benito, 2005; Gedye, 2010; Monereo, 2009; V. M. López, 2012; Sanmartí, 2007). Desde esta premisa, emerge la posibilidad de llevar a cabo un empleo estratégico de la misma con fines didácticos y pedagógicos (Gibbs, 2003). “La evaluación, bien comprendida y utilizada, es un aliado estratégico e insustituible para mejorar los procesos de enseñanza y aprendizaje” (Román & Murillo, 2009, p. 6). Gibbs y Simpson (2009) sostienen a este respecto que, ante la aproximación estratégica que llevan a cabo las/os estudiantes en su propio aprendizaje, el profesorado debe recurrir a un uso igualmente estratégico de la evaluación.

El valor de la evaluación como recurso didáctico estratégico reposa sobre los efectos o repercusiones que generalmente suscita en agentes evaluados y evaluadores. Revisada y reelaborada la propuesta inicial de Barberá (1999), advertimos siete grandes efectos ordenados en dos bloques: los relativos al poder de influencia de la evaluación y los vinculados a la información y el control que su uso genera.

- **Efectos de influencia.** Son aquellos que explican la capacidad de la evaluación para condicionar las actuaciones y representaciones de las principales audiencias involucradas en ella, en especial, de docentes y estudiantes. Estos efectos están

⁹ Estas diferencias y su efecto en la emergencia de conceptualizaciones diferentes será tratada en detalle a lo largo del próximo capítulo.

ligados entre sí y pueden intervenir en sentido favorable o desfavorable a los fines previstos según se manifiesten:

- ✓ *Efecto de transparencia.* Independientemente de lo que se detalle o declare explícitamente, “la evaluación pone de manifiesto todas nuestras concepciones docentes: lo que es la universidad, la naturaleza del proceso de enseñanza, el papel del docente, la relación profesor/alumno, etc.” (Santos, 1999, p. 370). A menudo implícitas e inconscientes, saca a la luz las creencias e intenciones educativas profundas (Barberá, 1999; Biggs, 2005; Jessop, 2012). Monereo y Castelló (2009b) atribuyen a la evaluación un efecto retroactivo sobre la enseñanza en este sentido. La literatura específica proporciona varios titulares que así lo corroboran: dime cómo evalúas y te diré qué y cómo enseñas (Sanmartí, 2007; Santos, 1999), y qué y cómo aprenden tus estudiantes (Monereo & Lemus, 2010; Sanmartí, 2007).
- ✓ *Efecto de directividad.* Unido al efecto anterior, la evaluación dirige la atención, el esfuerzo y el trabajo, priorizando un determinado estilo de aprendizaje frente a otros (Barberá, 1999; Meyer & Watson, 1991; Monroy & Hernández, 2014; Navaridas, 2002; Stobart, 2010). Este efecto tiene su máxima expresión en los instrumentos y técnicas evaluativas empleadas que, según sean interpretadas por el alumnado, favorecen unas estrategias de trabajo y aprendizaje u otras (Struyven, Dochy, & Janssens, 2005). Biggs (2005) denomina repercusión a la capacidad de la evaluación para determinar el aprendizaje por encima de lo establecido en el currículum oficial.
- ✓ *Efecto de regulación.* La evaluación es una herramienta de regulación en la medida en la que condiciona de forma evidente el trabajo y la actuación de docentes y discentes (Meyer, 1992a; B. Salinas & Cotillas, 2007; Sanmartí, 2007). Como afirma Nieto (2000) “la evaluación condiciona el trabajo y aprendizaje de los estudiantes más que cualquier otro factor instructivo, de modo que un buen sistema de evaluación constituye la mejor garantía para un aprendizaje relevante de la asignatura” (p. 307). En consecuencia, da lugar a ciclos de concentración, dedicación y esfuerzo diferentes, según las exigencias evaluativas que se planteen (Jessop, El Hakim, & Gibbs, 2014). Camilloni (1998) habla de validez de proacción para calificar este poder de movilización selectiva y dirigida asociada a la evaluación. Su efecto no sólo funciona sobre el alumnado. También representa condición de peso en el trabajo docente, en especial, cuando la evaluación prevista es externa y

vinculada a la rendición de cuentas (Brown, 2003b; Coll & Onrubia, 2002; Shepard, 2000).

- ✓ *Efecto de poder.* Los efectos de directividad y regulación presentados hacen de la evaluación un instrumento de poder en manos de quien la ejerce (Barberá, 1999; Batalloso, 2004; Toledo, 2005). El uso estratégico de la evaluación, reside precisamente en el empleo consciente de dicho poder, pero en condiciones éticas y pedagógicas, lejos de toda sospecha de abuso o mal uso (Ortiz, 2006; Santos & Moreno, 2004).

● **Efectos de información y control.** Son los derivados del uso de la evaluación como herramienta de indagación, ya sea a efectos pedagógicos o acreditativos¹⁰.

- ✓ *Efecto de control.* La evaluación representa un mecanismo para la rendición de cuentas interna o externa (Pearson, Vyas, Sensale, & Kim, 2001). Empleada como recurso para determinar niveles de logro y rendimiento, ejerce naturalmente un efecto de control sobre los agentes evaluados (Barberá, 1999).
- ✓ *Efecto de información.* La evaluación proporciona datos y evidencias a las audiencias implicadas directa o indirectamente en el proceso formativo. Representa un recurso para el diálogo, la comprensión y la mejora (Santos, 1993) y ofrece un soporte privilegiado para desempeñar funciones orientadoras sobre el aprendizaje y/o la enseñanza.
- ✓ *Efecto de clasificación.* La evaluación ejerce un efecto evidente de selección y clasificación, en la medida en la que discrimina niveles de consecución en los productos analizados o los logros previstos (Barberá, 1999; Santos, 2003; J. Vera & Esteve, 2001).

Usar estratégicamente la evaluación significa emplearla deliberadamente para estimular el esfuerzo, el trabajo y el tipo de actuación que requiere el logro de los objetivos previstos. Implica el empleo consciente de los efectos enumerados, basado en la previsión de su influencia potencial en la calidad y la cantidad del trabajo del alumnado. J. Chappuis (2014) habla en este sentido de evaluar consciente o concienzudamente, haciendo del aprendizaje el foco de atención permanente. Con ello, se abre al profesorado una amplia gama de oportunidades didácticas que exigen capacitación y cualificación profesional. Junto a ellas, también se les suma la responsabilidad de garantizar que la superación de un

¹⁰ Dichas finalidades serán definidas en el segundo capítulo al presentar las funciones de la evaluación.

determinado módulo o asignatura exija un esfuerzo cognitivo coherente con los logros establecidos (Monereo & Lemus, 2010).

Desde esta perspectiva, procurar un balance efectivo entre las oportunidades formativas que brinda la evaluación y su ejercicio responsable ofrece un nuevo argumento para profundizar en la competencia seleccionada.

3.3 La evaluación del aprendizaje de las/os estudiantes como factor clave asociado a la calidad de la enseñanza universitaria.

De acuerdo con Pérez Pérez¹¹ (2004) la mejora del sistema didáctico universitario puede producirse por tres vías preferentes: la metodología docente, la dimensión relacional que se establece entre alumnado y profesorado y la evaluación de los aprendizajes que, a su entender, representa un crisol para las dos anteriores. La presente investigación opta por esta última alternativa, elige la competencia evaluativa como punto de partida para impulsar la mejora de la docencia universitaria y, con ella, la del aprendizaje discente.

Además de un recurso técnico de gran potencialidad didáctica, la evaluación del aprendizaje de las/os estudiantes constituye uno de los factores clave asociados a la calidad de la docencia universitaria (Black & Wiliam, 1998b; Bain, 2007; Birenbaum, 1996; Stiggins, 1999; Villa, 2008; Zabalza, 2003). Representa una competencia específica indispensable por la influencia que ejerce sobre la calidad del aprendizaje y también sobre la de la enseñanza, a cuyo perfeccionamiento se encuentra estrechamente unida (Boud & asociados, 2010; V. M López, 2009; Nieto, 2000; Villar, 2009).

En consecuencia, revisar y mejorar la evaluación del aprendizaje, no sólo contribuye al aprendizaje de las/os estudiantes, sino también a la propia calidad de la enseñanza del profesorado universitario.

La evaluación es un proceso en el cual el docente cuestiona sus propios conocimientos, recibe retroalimentación de sus pares y construye nuevas prácticas pedagógicas. Así, el análisis del trabajo de los estudiantes se constituye en un

¹¹ La referencia se realiza utilizando los dos apellidos del autor para poder diferenciarlo de Rafael Pérez Juste.

camino para la comprensión del proceso de aprendizaje en general y de cada estudiante en particular. (J. Rodríguez, 2008, p. 8).

Este doble efecto reposa en la dinámica natural que se produce entre los tres ejes centrales de la actividad docente: el aprendizaje, la enseñanza y la evaluación. Como indican O. Esteve (2010) o Alsina (2011), los tres constituyen procesos inseparables, en especial enseñanza y evaluación, que para Monereo & Castelló (2009b) representan las dos caras de una misma moneda. “Modificando la manera en que evaluamos aquello que aprenden nuestros alumnos tenemos la posibilidad de modificar lo que realmente aprenden y, consecutivamente, también tenemos la oportunidad de modificar el modo en que se enseña lo que aprenden” (Monereo, 2009, p. 15).

La evaluación del aprendizaje representa una competencia docente particularmente indicada para impactar en la mejora general de la calidad de la enseñanza, por la estrecha relación que mantiene con otras competencias docentes relevantes. Esta conexión resulta especialmente evidente con respecto a la planificación y la elección metodológica, junto a la evaluación, las dos con mayor representación en la muestra de perfiles profesionales analizada (tabla 5). El modelo de alineación constructivista de Biggs (2005) reafirma esa unión, al plantear que la evaluación debe guardar plena coherencia con los objetivos de aprendizaje perseguidos (planificación) y con las actividades desempeñadas para su logro (metodología). De forma más precisa:

- Por un lado, admitido que enseñanza y evaluación forman parte de un mismo todo, el diseño de los dispositivos evaluativos no debe transcurrir desligado de la planificación general de la enseñanza, ni viceversa (Assessment Resources at Hong Kong University AR@HKU, 2009c; Biggs & Tang, 2011; Combs, Gibson, Hays, Saly, & Wend, 2008). Por ese motivo Mateo (2008) presenta la elección de métodos de evaluación como una de las decisiones clave de la planificación docente.
- Por otro lado, la evaluación puede suponer también un impulso positivo para la selección y la renovación metodológica. Cuando los objetivos de aprendizaje y la evaluación se planifican en consonancia, incrementa la probabilidad de realizar una selección estratégica y consciente de las tareas y situaciones de aprendizaje que mejor pueden contribuir a su logro.

Sostenemos con Rué (2008) que esta alineación de objetivos, métodos y estrategias de evaluación resulta especialmente relevante en el enfoque de competencias impulsado desde el EEES:

En esta nueva perspectiva, el fin propuesto para la formación [...] se realiza mediante un proceso específico que no es aleatorio ni secundario. Así, a trabajar en equipo se aprende trabajando en equipos eficientes y eficaces. A observar, se aprende observando realidades distintas, comparándolas. Resolver problemas significa inventar soluciones y verificar si funcionan o no, o hasta dónde o en qué condiciones lo son. Por lo tanto, finalidad, método y evaluación de resultados deben seguir una misma lógica. (p. 17).

3.4 El papel de la evaluación en la consolidación del nuevo paradigma formativo de la educación superior.

Si la evaluación determina por encima de otras decisiones curriculares las estrategias de aprendizaje y enseñanza adoptadas, la consolidación del nuevo paradigma formativo no parece posible a menos que se planteen dispositivos evaluativos acordes con el mismo (Bilbao, 2014; Olmos & Rodríguez, 2008; Ureña et al., 2009; Yániz & Villardón, 2006). La probabilidad de que una innovación curricular llegue a tener efecto es escasa si no se acompaña de cambios que incidan en el modo de concebir la evaluación (Baartman, Bastiaens, Kirschner, & Van der Vleuten, 2007; Bonsón & Benito, 2005; Watts & García-Carbonell, 2006). Dochy, Segers y Diederick (2002) hablan de la profecía de disolución autoinfringida (o autodisuelta), en referencia a aquellas innovaciones educativas que fracasan y desaparecen por no ir acompañadas de modelos de evaluación coherentes que las consoliden.

[...] Dado que un cambio de mentalidad implica un nuevo modo de observar y comprender una realidad y, como consecuencia, un cambio en el comportamiento práctico, a buen seguro será necesario que se produzcan, en muchas ocasiones, procesos de desaprendizaje que pueden ser tanto o más costosos que los propios procesos de aprendizaje. Muchos de estos procesos están relacionados con la concepción evaluativa que se tiene en estos momentos en la universidad y que

necesariamente ha de cambiar hacia una evaluación que además de juzgar sea capaz de educar. (A. Calatayud, 2008, p. 2).

El reto de la convergencia “no puede quedar en un simple maquillaje de lo existente” (Rué, 2008, p. 1). Requiere de un importante y costoso cambio de cultura y de mentalidad por parte de los distintos agentes implicados (Ion & Cano, 2011; Mateo & Vlachopoulos, 2013). Además de evitar que el paradigma formativo europeo pase por la universidad sin echar raíces, otra razón para centrar la atención en la competencia evaluativa del profesorado es que el propio EEES plantea cambios directos sobre ella. Presentadas a lo largo del capítulo cuarto, las nuevas exigencias evaluativas representan un reto profesional relevante para las instituciones formativas en general y, en particular para el profesorado.

3.5 El riesgo de perpetuación de los modelos disfuncionales de evaluación del aprendizaje.

En la formación básica de maestras/os la competencia evaluativa del profesorado adquiere una significación especial que proporciona nuevos argumentos para profundizar en ella. De acuerdo con A. Medina et al. (2011), el profesorado universitario representa el principal artífice del liderazgo formativo en otras etapas del sistema educativo. En el profesorado de enseñanza secundaria este influjo se ejerce por medio del Máster de Educación, mientras que en Educación Infantil y Primaria la capacitación y la incidencia tienen lugar en títulos de grado específicos. En ambos casos, se genera un ciclo de reproducción y retroalimentación en torno a las concepciones evaluativas¹².

El punto de partida se encuentra en las representaciones que el profesorado universitario sostiene sobre el aprendizaje, la enseñanza y, en especial, sobre la evaluación, en cuya configuración es bien conocida la incidencia de las experiencias discentes previas (Biggs, 2005; Feixas, 2002, 2008; Monereo & Castelló, 2009b). Situados en el rol docente, tendemos a reproducir aquellos patrones o pautas de actuación interiorizadas de forma inconsciente por medio de la experiencia lo que, llevado al terreno de la evaluación, conduciría a perpetuar una larga lista de deficiencias (Hortigüela et al., 2015; V. M. López, 2008; Palacios & López, 2013). Esta reproducción sucesiva de patrones

¹² El capítulo tercero ofrece una interpretación más detallada del mismo al identificar las causas de disfuncionalidad en las prácticas evaluativas.

disfuncionales de evaluación, conduce a una cultura evaluativa escolar que muestra muchos de los rasgos patológicos diagnosticados en la universitaria (Santos, 2003; Tejedor, 1999). Explica a su vez que, ya desde Educación Primaria, el alumnado manifieste una visión negativa de evaluación (M. A. Calatayud, 2004) y que las/os estudiantes de magisterio comiencen su formación profesional con esa misma creencia (Schommer, Beuchat, & Hernández, 2012).

Desde esta perspectiva, mejorar las prácticas evaluativas del sistema educativo exige el cambio de las representaciones que sostienen sobre la evaluación los distintos agentes involucrados (Bonsón & Benito, 2005; Knight, 2005; Monereo & Castelló, 2009b; Serafini, 2000), en especial el profesorado. Sin embargo, como señalan Buendía et al. (1999), el carácter encarnado de las concepciones educativas profundas, provoca que sean muy resistentes al cambio. Fuertemente enlazadas con las vivencias personales, su modificación resulta poco probable por la mera exposición a modelos y referentes teóricos, y así se viene haciendo tradicionalmente en facultades de educación y escuelas de magisterio (Marín & Cervi, 2006; Pozo et al., 2006).

Las experiencias evaluativas propiciadas directamente al alumnado en los grados educativos y en el Máster de Educación condicionan su praxis evaluativa por encima del discurso teórico (Santos, 2009; Smitter & Siso, 2007). En consecuencia, resulta prioritario modificar tales prácticas para no perpetuar las deficiencias actuales, ni propagarlas por efecto de reproducción en los niveles escolares y la enseñanza secundaria. Como concluye el próximo apartado, nuestra investigación se desarrolla específicamente en el contexto de la formación básica de maestras/os de Educación Infantil y Primaria.

3.6 Conclusión: investigación sobre la competencia evaluativa del profesorado en la formación básica de maestras/os y ciclo potencial de mejora.

El recorrido efectuado ofrece toda una serie de motivos interrelacionados que aconsejan articular esfuerzos efectivos por contribuir a la mejora de los procesos de evaluación del aprendizaje en la universidad. En consecuencia, el centro de interés del presente trabajo se sitúa en la competencia evaluativa del profesorado universitario, tal y como ha sido delimitada.

De forma más precisa, el estudio se focaliza en la competencia evaluativa de una sección del profesorado en particular, aquella vinculada a los grados asociados a la formación básica de maestras/os de Educación Infantil y Primaria. Conscientes del ciclo de influencias que se genera por la reproducción de los patrones o concepciones educativas profundas, advertimos en estas titulaciones un enclave estratégico para incidir en la mejora general del sistema educativo a largo plazo.

Como será desarrollado en próximos capítulos, avanzar hacia modelos de evaluación del aprendizaje de mayor calidad exige la revisión y el replanteamiento exhaustivo de las prácticas evaluativas y las concepciones profundas que subyacen a ellas, tanto en el papel que desarrolla el alumnado, como en el que concierne de forma directa al profesorado. Sin desestimar la influencia y el papel que juegan el resto de las audiencias implicadas (fundamentalmente, agentes empleadores, administradores, responsables institucionales, formadores o representantes sociales en general), la presente investigación parte del especial protagonismo de ambos colectivos.

Por relevante que sea el papel y la influencia del alumnado en la evaluación, la responsabilidad profesional recae formalmente sobre el profesorado (Bonsón, 2009; L. Prieto, 2004). En consecuencia, consideramos que el cambio en las concepciones y prácticas evaluativas debe iniciarse desde este colectivo, atendiendo a la competencia evaluativa que manifiesta en sus decisiones e intervenciones. Las actuaciones y creencias de las/os estudiantes con respecto a la evaluación forman parte de la responsabilidad pedagógica del profesorado (Poblete, 2007), de modo que debería orientar en las mismas una evolución deseable mediante su propia práctica.

La precisión y explicitación de los fines de la acción educativa, pasa a ser una herramienta indispensable para guiar la acción educativa. Es por ello que el docente, formador de profesores, tiene la responsabilidad de que su metodología refleje y esté en sintonía con lo propuesto por la reforma educacional. (Schommer et al., p. 473).

Desarrollar esta competencia entre el profesorado universitario, dando lugar a nuevas experiencias y perspectivas evaluativas, supondría una alternativa para incidir en la visión del alumnado desde la práctica. I. Álvarez (2008a) considera que las/os docentes deben avanzar en la cultura de la evaluación como oportunidad de aprendizaje y, a la vez, introducir al alumnado en esa misma cultura mediante experiencias directas que contribuyan a modificar sus representaciones iniciales y sus concepciones futuras.

En las titulaciones vinculadas a la formación básica de maestras/os, dicha responsabilidad adopta un calado más profundo. En ellas, el profesorado se enfrenta al doble reto de evaluar el aprendizaje de las/os estudiantes y de promover el desarrollo de su competencia evaluativa como futuras/os profesionales de la enseñanza (Bilbao, 2010).

Sumando las consideraciones anteriores, sostenemos que incidir en la competencia evaluativa del profesorado vinculado a la formación básica de maestras/os representa un foco estratégico de acceso a la mejora general de la evaluación en el sistema educativo. La figura 7 representa la secuencia de desarrollo progresivo que tenemos en mente, y a la que este trabajo sumaría una pequeña contribución inicial:

- El desarrollo de la competencia evaluativa del profesorado universitario, debería ocasionar, en primer término, nuevas prácticas y concepciones evaluativas dentro del propio colectivo docente. Manifestadas en la práctica...
- ...las nuevas actuaciones del profesorado favorecerían en el alumnado experiencias alternativas con respecto a la evaluación del aprendizaje y, con ello, nuevas concepciones sobre la misma (V. M. López, 2008). Con ello, comenzaría a perder vigencia esa visión negativa y reduccionista de la evaluación que la define como una acción sumativa, punitiva y finalista. Por efecto de reproducción de vivencias...
- ...estas nuevas concepciones y prácticas evaluativas experimentadas por el alumnado en primera persona podrían trasladarse, a medio plazo, a los escenarios formativos en los que vaya a tener lugar su ejercicio profesional como maestras/os. Ello favorecería la difusión de una imagen renovada y más pedagógica del fenómeno evaluativo desde los niveles inferiores del sistema educativo. Para que este efecto se extienda también a la enseñanza secundaria, sería necesario abordar específicamente el desarrollo de la competencia evaluativa del profesorado al cargo del Máster de Educación, aunque ello no es objeto del presente trabajo. De este modo...
- ..., y sin desestimar los procesos de influencia e interferencia mutua que puedan tener lugar, las nuevas formas de ejercer y entender la evaluación podrían extenderse paulatinamente al resto de audiencias involucradas (agentes empleadores, administradores, etc.), dando lugar a la configuración de una nueva cultura evaluativa a largo plazo.

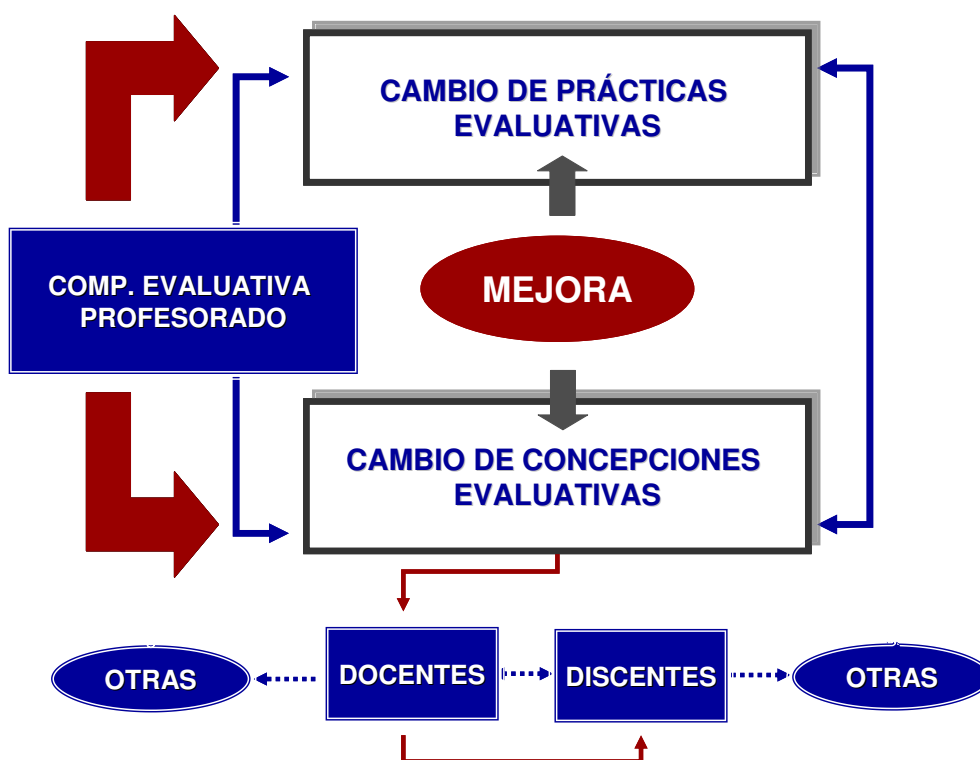


Figura 7. La competencia evaluativa del profesorado universitario de las titulaciones dedicadas a la formación de maestros como elemento estratégico para la mejora.

En síntesis, como indica Fullan (2007) toda modificación que afecte al terreno evaluativo es difícil, lenta y compleja, porque exige realizar adaptaciones relevantes en la enseñanza y el replanteamiento de las convicciones profesionales y personales sobre las que ésta se apoya.

4. El capítulo a grandes rasgos

La configuración del EEES ha traído consigo una serie de cambios pedagógicos y estructurales a las universidades de los estados miembro. Su consolidación progresiva ha modificado el contexto formativo y, con él, el papel atribuido al profesorado universitario. De acuerdo a los nuevos tiempos y las nuevas pretensiones, tiene lugar una ampliación y redefinición de sus principales roles o espacios de actividad, en especial, de la docencia. Retomando el clásico debate sobre los enfoques docentes, se apuesta por un modelo general de enseñanza orientado al desarrollo de competencias profesionales en el alumnado (aprendizaje).

La redefinición del cometido profesional del profesorado universitario ha dado lugar en los últimos años a la emergencia de propuestas de sistematización elaboradas desde perspectivas y premisas diferentes. No existe un único perfil profesional, pero las competencias genéricas y específicas integradas en las diversas aproximaciones, configuran una visión general de la enseñanza acorde con el planteamiento de la convergencia europea.

La evaluación del aprendizaje discente es una de las competencias específicas con mayor presencia entre los perfiles analizados y la presente investigación centra su atención en ella. Fruto de la movilización armónica e integrada de conocimientos declarativos, funcionales y de actitudes y predisposiciones, la competencia evaluativa se concreta en la respuesta eficaz y responsable del profesorado a los dilemas y demandas que plantea la evaluación del aprendizaje del alumnado en el contexto universitario. Se trata de una competencia versátil y multifuncional, asociada específicamente al rol docente del profesorado y abierta a su mejora y reconstrucción permanente.

Investigar sobre la competencia evaluativa del profesorado puede contribuir a solventar parte de las clásicas deficiencias apegadas a los sistemas de evaluación universitarios, revertir en la mejora del aprendizaje y de la enseñanza, y contribuir al reto pendiente de la consolidación del paradigma formativo del EEES. Resulta especialmente pertinente en las titulaciones vinculadas a la formación básica de maestras/os de Educación Infantil y Primaria, donde la evaluación figura entre las competencias específicas a desarrollar por el alumnado. Por efecto de reproducción, las experiencias evaluativas vivenciadas en este contexto, tienen alta probabilidad de ser difundidas en los niveles educativos de destino por las/os estudiantes actuales (futuras/os evaluadoras/es) finalizados sus estudios.

Por todo lo anterior, consideramos que contribuir al desarrollo de la competencia evaluativa del profesorado en los grados vinculados a la formación profesional de maestras/os representa una medida estratégica de mejora. Su efecto mediaría a dos niveles complementarios: a nivel directo, favoreciendo una mejor cualificación profesional docente y preparación de las futuras generaciones de maestras/os, y a nivel indirecto, contando con el impacto que pueda ejercer en la calidad del sistema educativo general la incorporación de profesionales con una visión renovada de la evaluación del aprendizaje.

Emprender la línea de actuación trazada exige, en primer término, delimitar un marco de referencia común que haga posible el entendimiento mutuo. El próximo capítulo responde a esta necesidad de clarificación terminológica y conceptual.

CAPÍTULO 2.

CONCEPTUALIZACIÓN: ESTABLECIMIENTO DE UN MARCO COMÚN EN TERMINOLOGÍA EVALUATIVA

Aunque representa un objeto de estudio ampliamente recurrido y desarrollado, la evaluación del aprendizaje suscita confusión y una ambigüedad terminológica evidente en la investigación, en la literatura y, por extensión, en las propias instituciones formativas (B. Álvarez et al., 2007; J. M. Álvarez, 2003; Joint Committee on Standards for Educational Evaluation-JCSEE, 2007; V. M. López, 2012; V. M. López et al., 2007; Nevo, 1997). Esta divergencia de interpretaciones genera serias dificultades para consensuar y difundir vías de mejora, de modo que la delimitación de glosarios que permitan el entendimiento mutuo representa un recurso cada vez más extendido¹³.

En estas circunstancias, profundizar sobre la competencia evaluativa docente exige llevar a cabo una conceptualización y clarificación terminológica previa, que explicita el sentido y el significado atribuido a la evaluación del aprendizaje. A ello responde este segundo capítulo, que incorpora seis apartados en su desarrollo y una síntesis final de sus principales conclusiones:

- El primer apartado revisa la evolución que ha sufrido el término evaluación en los últimos años, delimitando los diferentes paradigmas y enfoques surgidos en su interpretación y su impacto en la situación general de confusión reseñada.
- Identificadas las posiciones de partida, el segundo apartado presenta la definición de evaluación por la que aboga la presente investigación.

¹³ Partiendo de la importante contribución inicial del tesoro de evaluación de Scriven (1991), así lo hemos podido constatar en los espacios web de los servicios y unidades de evaluación de diversas universidades. A modo de ejemplo, algunos de los considerados en la presente investigación son los elaborados por el Assessment Resources de la Universidad de Hong Kong (AR@HKU, 2009b), el American Public University System (APUS, 2013), la Assessment Office de la Universidad de Hawai'i Manoa (UH Manoa, 2013), el Centre for Assessment & Research Studies (CARS, 2003) o el del Outcomes Assessment Committee del Nevada State College (NSC, 2009).

- El tercer apartado expone las cuatro fases en las que se desarrolla el proceso evaluativo y la secuencia de pasos o tareas integrada en cada una de ellas.
- El cuarto apartado se detiene en la exploración de las funciones sociales y pedagógicas atribuidas a la evaluación del aprendizaje.
- El apartado quinto plantea una clasificación revisada de las principales tipologías presentes en la literatura al explorar la variedad de formas y modalidades en la evaluación del aprendizaje.
- El sexto apartado realiza una breve introducción a la reflexión sobre la calidad evaluativa, desarrollado en profundidad a lo largo del capítulo cuarto.
- El capítulo concluye con una síntesis que cierra los principales aspectos abordados en su desarrollo y delimita las posiciones de partida.

Para completar este recorrido, el **ANEXO 02** ofrece un pequeño glosario de la terminología evaluativa más empleada a lo largo de la presente investigación.

1. La evaluación del aprendizaje y su evolución histórica: distintas interpretaciones en la ciudad de Babel

Pero el Señor bajó para observar la ciudad y la torre que los hombres estaban construyendo, y se dijo: «Todos forman un solo pueblo y hablan un solo idioma; esto es sólo el comienzo de sus obras, y todo lo que se propongan lo podrán lograr. Será mejor que bajemos a confundir su idioma, para que ya no se entiendan entre ellos mismos».

De esta manera el Señor los dispersó desde allí por toda la tierra, y por lo tanto dejaron de construir la ciudad. Por eso a la ciudad se le llamó Babel, porque fue allí donde el Señor confundió el idioma de toda la gente de la tierra, y de donde los dispersó por todo el mundo.

La torre de Babel, Génesis 11:1-9.

La complejidad natural de la evaluación y la diversidad de objetos, audiencias y propósitos a los que sirve la hacen especialmente proclive a la polisemia y la ambigüedad. En consecuencia, el reflejo que devuelve de ella la investigación y la literatura específica guarda cierta similitud con la popular fábula de Babel. Plantea Zabalza (2011) que algunos aspectos relacionados con la docencia terminan en una jungla semántica, en la que la obsesión por incorporar matices conduce a nuevos términos que confunden más que clarifican. Algo así parece haber ocurrido en la evaluación del aprendizaje, donde la variedad de acepciones parece más un síntoma de confusión que de riqueza (Birenbaum, 1996). La proliferación de estudios y contribuciones de diverso tipo no sólo no han resuelto esta clásica dificultad, sino que parece haber acrecentado la emergencia de nuevos términos e interpretaciones.

La conclusión es a todas luces evidente: cuando nos comunicamos, el profesorado universitario, utilizamos los mismos términos con significados diferentes, lo que nos conduce a una mala interpretación de la información y a una casi imposible tarea de construir y elaborar nuestro conocimiento profesional. En definitiva, es preciso que acordemos el significado de los términos que utilizamos en el contexto de la evaluación ya que ello nos permitirá poder comunicarnos de una forma más eficiente pero, sobre todo, poder ir construyendo un conocimiento compartido sobre nuestras prácticas evaluativas. (G. Rodríguez & Ibarra, 2012, p. 152).

Uno de los factores asociados a esta polisemia tiene que ver con la evolución histórica que ha sufrido el término evaluación a lo largo de las últimas décadas. Su origen se remonta al menos al año 2000 a.C., aunque la trayectoria de la evaluación como actividad profesional no arranca hasta mediados del siglo XIX (Stufflebeam & Shinkfield, 1987). Desde ese momento, se han sucedido distintas maneras de interpretar el fenómeno evaluativo en una evolución que, según House (1993), se ha traducido en cambios conceptuales, estructurales, metodológicos y prácticos. La evaluación del aprendizaje, como una de las muchas manifestaciones que puede tener el fenómeno evaluativo, recoge sobre sí los rasgos y repercusiones de dicha evolución.

Guba y Lincoln (1989) realizaron hace más de dos décadas un trabajo de síntesis ampliamente referido. En él identificaron cuatro grandes generaciones en el modo de realizar la evaluación: la de la medida, la de la descripción, la del juicio y la respondiente-constructivista. En conjunto, representan cuatro perspectivas que se suceden desde una

lógica de complementariedad, dando lugar a la cantidad y variedad de propuestas de definición que se constatan en la actualidad.

“Ninguna introducción a la evaluación como práctica profesional estará completa si no se presta la debida atención a la evolución histórica del tema” (Stufflebeam & Shinkfield, 1987, p. 32). A tal efecto, la tabla 6 recupera los principales rasgos de la visión evaluativa sostenida en cada generación.

Tabla 6

Principales rasgos de las cuatro generaciones de la evaluación (Guba & Lincoln, 1989)

a) EVALUACIÓN COMO MEDIDA

- Tiene lugar a partir de la segunda mitad del siglo XIX principalmente, cuando las prácticas evaluativas comienzan a sistematizarse. A menudo se denomina como fase pre-Tyleriana, en contraste con la generación evaluativa que le sucede.
- En este momento la evaluación se concibe como medición. Se orienta a la obtención de medidas tipificadas, contagiada del interés científico por la medición de las conductas humanas. Las contribuciones de Binet y Galton en materia psicométrica son particularmente representativas en este sentido.
- La evaluación se centra en los agentes evaluados y los instrumentos. El papel del agente evaluador es el del técnico experto en la aplicación de instrumentos de evaluación, tipificados para la medición de uno o varios atributos.

b) EVALUACIÓN COMO DESCRIPCIÓN

- Las limitaciones de la generación anterior dan lugar a una nueva visión de la evaluación como descripción. Representa un hito histórico de máxima relevancia porque se corresponde con surgimiento de la evaluación educativa, a partir de las contribuciones de una línea de trabajo iniciada por Tyler en la década de los treinta y publicada años más tarde (Tyler, 1950).
- En esta generación se entiende el proceso educativo como un todo. Admitido que la evaluación no sólo debe centrarse en el alumnado, surgen nuevos focos de interés, entre los que destacan el currículo y los programas formativos, especialmente sus objetivos.
- La evaluación se concibe desde una perspectiva más dinámica y pasa a constituir un mecanismo para la mejora continua del currículo y la institución formativa. En consecuencia, se centra en la descripción de oportunidades y limitaciones con respecto a los objetivos fijados, de posibilidades y limitaciones para la mejora formativa.
- Desde esta perspectiva el rol del evaluador se amplía: a la medición se le añade la descripción como medio para contribuir a la mejora.
- Esta etapa se extiende hasta la década de los sesenta, atravesando distintas fases de desarrollo. Su comienzo está determinado por la aportación que supone el modelo de evaluación educativa de Tyler, que se extiende progresivamente en las dos primeras décadas. Los años cincuenta se caracterizan por la realización de inversiones económicas para reforzar la función evaluativa, aunque se descuida su orientación e impacto en la mejora. En consecuencia, en la década de los sesenta se generaliza el desánimo y se detectan las principales limitaciones que darían lugar a una nueva generación:
 - ✓ Cronbach (1963) señaló la necesidad de vincular la evaluación a la toma de decisión. También puso de manifiesto dudas sobre su utilidad y relevancia, advirtiendo la necesidad de analizar la pertinencia y adecuación de los propios criterios u objetivos formativos.
 - ✓ Scriven (1967) criticó el emplazamiento final de la evaluación por limitar su potencialidad para incidir en la mejora y en el logro de los objetivos establecidos.

c) EVALUACIÓN COMO JUICIO

- Surge como reacción a las limitaciones advertidas en el tramo final de la generación anterior y en respuesta a la crisis de identidad de las/os profesionales de la evaluación. Su desarrollo se extiende en las décadas de los setenta y ochenta.
- De la descripción se evoluciona al juicio, donde los logros evidenciados son sometidos a valoración de acuerdo con unos criterios previamente establecidos. En consecuencia, el agente evaluador se convierte en un juez, sin renunciar a las funciones técnicas y descriptivas.
- Es un momento de eclosión caracterizado por el desarrollo de fundamentos para la profesionalización de la actividad evaluadora (programas formativos, asociaciones y centros especializados, publicaciones, etc.). Con ello, emerge también la inquietud por la calidad de las evaluaciones, dando lugar a la metaevaluación.
- En este caldo de cultivo surge una amplia variedad de modelos evaluativos con gran riqueza de matices. Entre ellos se distinguen dos grandes bloques con rasgos conceptuales y metodológicos propios. Algunas propuestas hablan de modelos cualitativos y cuantitativos, mientras que otras se distinguen modelos clásicos y alternativos.
- Los modelos clásicos se desarrollaron durante la década de los setenta básicamente. Comparten la lógica racional cuantitativa, aunque responden a diversas tendencias. Entre ellos resultan especialmente destacables el modelo CIPP (Contextos, Información, Proceso y Producto) de Stufflebeam (1971), el modelo CSE (Centre for the Study of Evaluation) de Alkin (1969), así como las primeras contribuciones de Stake (1967). En línea con las recomendaciones de Cronbach (1963), la evaluación aparece vinculada a la toma de decisión en todos ellos.
- Stufflebeam y Shinkfield (1987) proponen una clasificación más detallada. Parten de la premisa de que sin juicio, no existe evaluación verdadera; en consecuencia, califican de pseudoevaluaciones o cuasievaluaciones las prácticas evaluativas de generaciones anteriores. Dentro de las evaluaciones verdaderas, discriminan cuatro tipos de estudio dependiendo del nivel al que vayan a ser adoptadas las decisiones resultantes:
 - ✓ Orientados a la decisión, que se dirigen principalmente a los niveles administrativos (por ejemplo, modelo CIPP de Stufflebeam, 1971; modelo CSE de Alkin, 1969).
 - ✓ Centrados en el cliente, que se orientan a quienes proporcionan el servicio, esto es, al personal (entre ellos, modelo respondiente de Stake, 1975, 1975/2011; B. McDonald, 2008).
 - ✓ Políticos, destinados a apoyar decisiones políticas como su propio nombre indica (Owens, 1971; cit. en Stufflebeam & Shinkfield, 1987).
 - ✓ Basados en el consumidor, que se dirigen a la perspectiva de los usuarios o receptores del servicio (Scriven, 1967).

d) EVALUACIÓN RESPONDENTE-CONSTRUCTIVISTA

- La cuarta generación evaluativa emerge a finales de la década de los ochenta, como resultado del aprendizaje construido de las tres anteriores. Recibe el nombre de respondiente-constructivista en alusión a sus dos grandes fuentes de influencia. Del modelo de evaluación respondiente de Stake (1975, 1975/2011) toma la orientación a los responsables y del paradigma constructivista, los planteamientos metodológicos y la sensibilidad con la construcción de significados.
- El rol evaluador se amplía, sin renunciar a los atributos de las generaciones anteriores. Pasa de juez a investigador, capaz de construir hechos que le permitan comprender el acto de aprender y, por lo tanto, de valorarlo. Entre sus rasgos fundamentales se cuentan:
 - ✓ Los resultados de las evaluaciones no son descripciones, sino construcciones generadas por agentes evaluadores y evaluados en un proceso interactivo.
 - ✓ Dichas construcciones se llevan a cabo desde un determinado sistema de valores y se reconocen influenciadas por el contexto de referencia.
 - ✓ El proceso de evaluación implica la interacción los agentes evaluadores y los evaluados. Cabe la negociación sobre el desarrollo de la evaluación y el uso de sus resultados. En consecuencia, esta generación también se conoce como la de la negociación.

- En síntesis, la cuarta generación dibuja un proceso evaluativo más complejo que las anteriores. Fruto de una actividad investigadora, la evaluación se convierte en un generador de realidad, cuyos resultados no pueden predecirse. La construcción se efectúa por medio de procesos de interacción e intercambio entre los agentes implicados y es permeable al contexto.

e) EVALUACIÓN PARA LA CALIDAD

Nota: Reelaboración propia a partir de diversas fuentes (Carmena, Asensio, De la Cuesta, Doncel, & Rodríguez, 2005; T. Escudero, 2003; Guba & Lincoln, 1989; R. López, 2001; Madaus & Stufflebeam, 1984; Mateo, 2000; Monedero, 1998; Muñoz, 2007; Stufflebeam & Shinkfield, 1987). La celda sombreada señala la posición de partida adoptada en el presente estudio.

Siguiendo a T. Escudero (2003) y Smitter y Siso (2007), las prácticas evaluativas actuales son propias de la cuarta generación, aunque hay voces que anuncian la próxima emergencia de una quinta. Muñoz (2007) describe en este sentido la generación de la evaluación para la calidad, caracterizada por hacer del proceso evaluativo un recurso para la formación permanente:

La evaluación para la calidad se da cuando los participantes del proceso (evaluador-evaluado), además de evaluar, miden, ponderan, acreditan, valoran y acrecientan.

Los actores serán capaces de reconocer sus fortalezas, pero también sus debilidades y buscarán la opción para crecer. Aquí la evaluación va más allá de un tú o un yo, es un `nosotros´. (p. 189).

Las cuatro generaciones recorridas, ofrecen una representación simplificada de las prácticas evaluativas más extendidas y generalizadas en cada momento. Esta delimitación guarda una estrecha conexión con dos conceptos clave que serán abordados en los próximos apartados y que, a menudo, aparecen entremezclados por su interdependencia (ver figura 8).

- Los paradigmas de evaluación, que son sistemas de creencias complejos, que motivan el comportamiento evaluativo propio en cada generación¹⁴. Se discriminan cuatro paradigmas principales: la medida, el procedimiento, la indagación y el control de calidad.
- Los enfoques de evaluación, que representan las principales tendencias de actuación que se constatan en las distintas generaciones. Reposan sobre una serie

¹⁴ La vinculación de ambos conceptos es tal que, la presentación efectuada para cada generación, incorpora referencias a las convicciones y creencias de partida, sin las cuales, su comprensión no sería posible (Tabla 6).

de creencias y supuestos previos, determinados por los paradigmas. Ofrecen la categorización más amplia y vaga al clasificar en dos bloques la amplia variedad de propuestas existentes: el enfoque evaluativo tradicional y el alternativo.

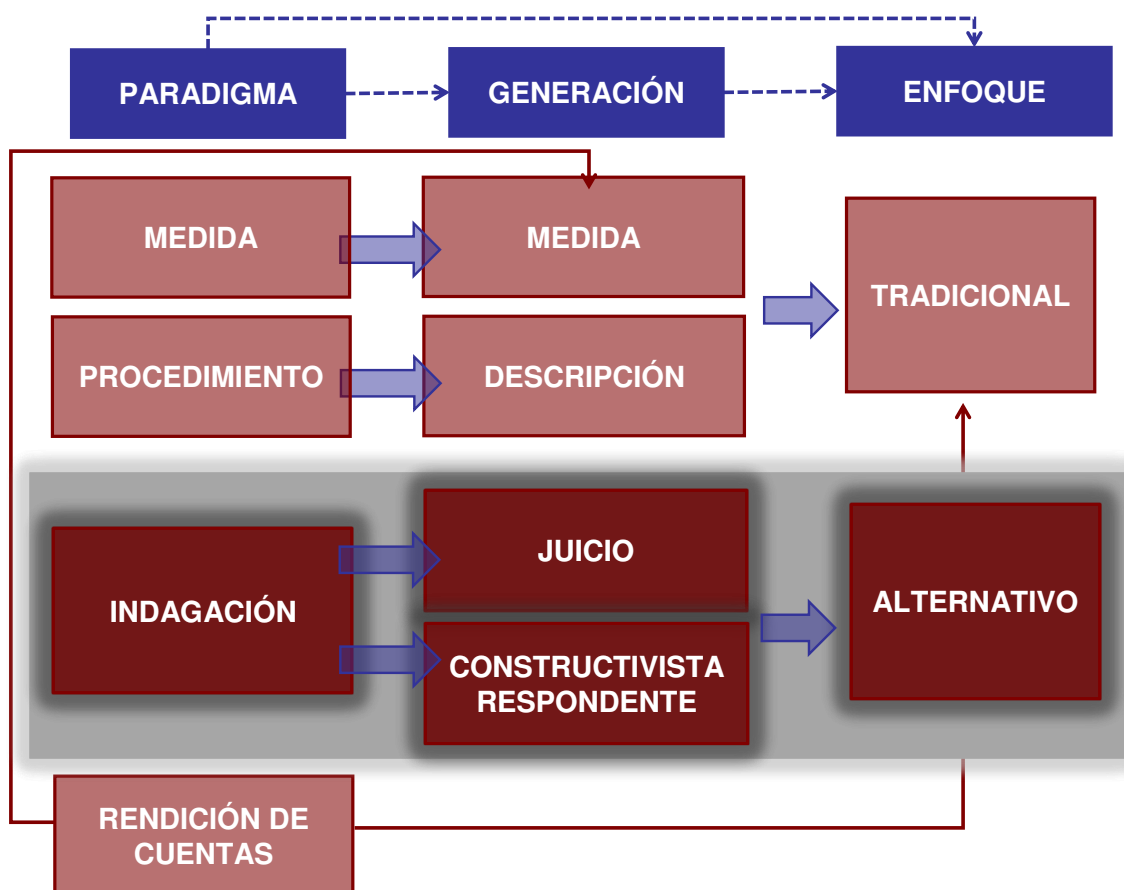


Figura 8. Las generaciones evolutivas del concepto evaluación y su relación con los paradigmas de evaluación y los enfoques evaluativos.

1.1 Paradigmas de evaluación.

El recorrido revisado en el apartado anterior revela cuatro grandes representaciones de la práctica evaluativa, elaboradas desde diferentes paradigmas de la evaluación. Adecuando la definición que plantean Lincoln y Guba (1985) para la investigación, entendemos los paradigmas evaluativos como sistemas de creencias complejos sobre la evaluación y sus componentes fundamentales (principalmente, sobre sus propósitos, la naturaleza de los objetos a los que se dirige, el cometido del agente evaluador y evaluado y la relación entre ambos, y los procedimientos de recopilación y análisis aplicables).

Partiendo del análisis de las distintas concepciones desarrolladas a lo largo de su evolución, Serafini (2000) identificó tres grandes paradigmas evaluativos a comienzos de la década pasada. Pocos años más tarde, Falchikov (2005) completó la clasificación añadiendo uno más. Descritos en próximas líneas, cada uno de los cuatro paradigmas devuelve una determinada visión de la evaluación, como medida, procedimiento, indagación o control de calidad. Como en la presente investigación, ambos trabajos están desarrollados desde la perspectiva educativa y se centran prioritariamente en la evaluación del aprendizaje. En consecuencia, las alusiones generales a agentes evaluadores y evaluados se sustituyen directamente por referencias a alumnado y profesorado.

El primer paradigma concibe la *evaluación como medida* y se encuentra estrechamente relacionado con la visión positivista de la realidad y del conocimiento. En esta aproximación a la evaluación, el aprendiz se considera un recipiente vacío y la instrucción transmisiva la vía preferente para llenarlo (Birenbaum, 1996). “En el paradigma de la evaluación como medida, se considera que el conocimiento existe con independencia del aprendiz, y los estudiantes trabajan para adquirirlo, no para construirlo” (Serafini, 2000, p. 385). En consecuencia, la evaluación se basa en el empleo de escalas y test estandarizados referidos a norma (Falchikov, 2005).

El paradigma de la evaluación como medida se orienta a la objetividad, la sistematicidad y la credibilidad o fiabilidad. En consecuencia, no existe una inquietud manifiesta por la implicación del alumnado y el profesorado en la evaluación. El rol docente es limitado y se restringe a la administración de pruebas diseñadas externamente, y puntuadas por medio de procedimientos computerizados. La participación en el diseño de la evaluación es prácticamente nula y tampoco es frecuente el empleo de los resultados generados. En la misma línea, la participación del alumnado se limita a la cumplimentación de las pruebas sin que apenas se le brinden oportunidades para la reflexión y la autoevaluación.

La evolución histórica del concepto revela un segundo paradigma que caracteriza la *evaluación como procedimiento*. Esta nueva visión mantiene los supuestos epistemológicos de la anterior y los propósitos fundamentales atribuidos a la evaluación. “En el paradigma de la evaluación como procedimiento se sigue solicitando al profesorado la medición objetiva de las habilidades de los estudiantes y el reporte de la información a audiencias externas en formatos numéricos” (Serafini, 2000, p. 386). El único cambio con respecto al anterior paradigma se centra en los procedimientos de recopilación de información empleados que, en este caso, son mucho más cualitativos (Falchikov, 2005).

Analizados sus rasgos generales, puede concluirse que los paradigmas de la evaluación como medición y como procedimiento subyacen a las prácticas evaluativas de las dos primeras generaciones (figura 8). Orientadas respectivamente a la medición y la descripción, ambas visiones se caracterizan por pretender la exploración de una realidad que se considera única, aunque emplean diferentes recursos y generan distinto tipo de información en dicho empeño. La generación evaluativa del juicio y la constructivista y respondiente se relaciona principalmente con el tercer paradigma o sistema de creencias.

El tercer paradigma representa un cambio relevante con respecto a las dos visiones precedentes y define la evaluación **como indagación o investigación**. A diferencia de lo que sucedía en el paradigma anterior, conlleva la adopción de nuevos supuestos epistemológicos, nuevos propósitos y nuevas audiencias, aunque se siguen aplicando los mismos procedimientos evaluativos. Para Falchikov (2005) y Serafini (2000) representa el paradigma evaluativo de referencia aunque, por el momento, su vigencia sea más una aspiración que una realidad.

Los nuevos supuestos epistemológicos provocan la superación de la interpretación positivista del aprendizaje. En su lugar se adopta la perspectiva constructivista y, en consecuencia, comienza a reconocerse el importante papel que desempeñan en él alumnado y profesorado (Birenbaum, 1996).

Como hace constar su propia denominación, la evaluación adopta la indagación y la comprensión como principal propósito. Se abre paso a un nuevo planteamiento evaluativo ideado para favorecer decisiones curriculares que fortalezcan el aprendizaje (Serafini, 2002). En consecuencia, la evaluación deja de ser un apéndice final y pasa a considerarse parte integrante del proceso formativo.

Aquí, el profesor usa varias técnicas de evaluación cualitativa y cuantitativa para indagar sobre aprendices particulares y su proceso de aprendizaje. Es un proceso de investigación e interpretación, empleado para promover la reflexión sobre los aprendizajes, actitudes y habilidades académicas de los estudiantes. (Serafini, 2000, p. 387).

Unido a todo ello, también se produce la redefinición de los principales roles dentro de la evaluación. Concebida como investigación o indagación, el foco de atención se desplaza de las audiencias externas a las internas, atendiendo preferentemente las necesidades de alumnado y profesorado. En consecuencia, el profesorado deja de ser un

mero administrador de pruebas estandarizadas e incrementan las expectativas de actuación y aprovechamiento sobre el alumnado.

Falchikov (2005) suma un cuarto paradigma a los tres anteriores y alerta de los riesgos que supondría la instalación definitiva en el mismo. Se trata del paradigma de la **evaluación como control de calidad**, derivado de las aportaciones de Pearson et al. (2001) sobre la rendición de cuentas.

Tomando como referencia el contexto educativo estadounidense, Pearson et al. (2001) detectaron una inercia pendular o de alternancia en el uso de modelos de rendición de cuentas externos e internos. En correspondencia con la primera generación evaluativa y el paradigma de la medición, su análisis sugiere para la rendición de cuentas una etapa inicial de control externo, basada en la aplicación de pruebas estandarizadas. Los riesgos y las repercusiones negativas de esta forma de proceder provocan una contra respuesta, dando lugar al desarrollo de procesos internos de rendición de cuentas a cargo de los propios centros y del profesorado. Esta segunda fase supone la aproximación al paradigma evaluativo de la indagación, propio de las generaciones tercera y cuarta. El desarrollo de este tipo de prácticas evaluativas internas representa una oportunidad de mejora destacable por la orientación específica y contextualizada que propicia. Sin embargo, también genera recelos y escepticismo entre quienes adoptan lineamientos evaluativos más apegados a la objetividad y la comparabilidad. En tales circunstancias, Falchikov (2005) advierte en este cuarto paradigma el retorno al planteamiento externo de la rendición de cuentas. Ello supondría tomar distancia de la indagación y retroceder a supuestos evaluativos próximos al paradigma de la medición.

La presente investigación se sitúa en el tercer paradigma evaluativo y caracteriza la evaluación como un proceso de indagación. Para llevar a cabo una mejor delimitación de la posición de partida, el siguiente apartado profundiza en las diferentes conceptualizadas que surgen dentro de dicha visión evaluativa.

1.1.1 La evaluación como indagación: revisión de conceptualizaciones.

Partiendo nuevamente de la perspectiva educativa, Falchikov (2005) discrimina seis conceptualizaciones distintas sobre la evaluación dentro del paradigma de la indagación. Las seis sostienen una visión constructivista del proceso aprendizaje, reconocen la participación del alumnado en el mismo e interpretan la comprensión como propósito central de la evaluación. Es habitual hallarlas referidas de forma indistinta, aunque cada una de

ellas incorpora matices específicos y una posición particular dentro del mismo sistema de creencias evaluativas.

La tabla 7 muestra la reinterpretación de la propuesta inicial de Falchikov (2005), incorporando ciertos matices para hacerla extensible al reflejo y desarrollo que ha tenido en la literatura. Aunque su presentación se realice de forma independiente, las distintas conceptualizaciones no son excluyentes entre sí y, a menudo se manifiestan solapadas y combinadas. De hecho, el trabajo presentado adopta la primera conceptualización como referencia principal dentro del paradigma e integra en su desarrollo las cinco restantes.

Tabla 7

Principales conceptualizaciones dentro del paradigma de la indagación: reinterpretación de la propuesta de Falchikov (2005)

PARADIGMA TERCERO: EVALUACIÓN COMO INDAGACIÓN
<i>Evaluación como aprendizaje o evaluación para el aprendizaje.</i> • De forma general, se caracteriza por considerar que la evaluación forma parte del propio proceso de instrucción y que puede emplearse para promover y apoyar el aprendizaje, no sólo para determinar calificaciones y rendir cuentas (Assessment Reform Group-ARG, 1999; S. Chappuis & Stiggins, 2002; Stiggins, 2005b). • Prioriza la función pedagógica de la evaluación (formadora y formativa) frente a la acreditativa (sumativa), según la conceptualización establecida en un próximo apartado. • Propone un uso consciente y deliberado de la evaluación como recurso para la promoción del aprendizaje (J. Chappuis, 2014; Stobart, 2010). Apoyada en las orientaciones específicas que se desprenden de la investigación y la literatura, esta conceptualización dirige la praxis evaluativa de acuerdo a una serie de principios con impacto potencial en el aprendizaje. Entre los más relevantes se encuentran: el énfasis de la evaluación en el aprendizaje, la atención al impacto emocional de la evaluación, la interiorización de criterios y condiciones por parte del alumnado, la participación de las/os estudiantes en el proceso evaluativo y la guía constructiva del aprendizaje por medio del feedback y la proacción.
<i>Evaluación autónoma o participativa</i> • Se caracteriza por plantear la redefinición de la participación del alumnado en la evaluación y su grado de poder y control. El principal supuesto departida es que la implicación discente en este proceso representa una oportunidad didáctica y la respuesta a muchas de las clásicas deficiencias advertidas en el mismo (I. Álvarez, 2008a; A. Brew, 2003; Boud & Falchikov, 2007; Falchikov & Goldfinch, 2000). • La conceptualización autónoma o participativa presenta la intervención de las/os estudiantes en la evaluación como una experiencia de aprendizaje. Como será expuesto al detallar las principales tipologías evaluativas (tercer apartado del presente capítulo), esta participación puede tener lugar en modalidades organizativas diferentes (autoevaluación, heteroevaluación o coevaluación), en distintas fases del proceso (planificación, implementación, establecimiento de conclusiones y/o revisión) y con desigual nivel de incidencia (con repercusiones en la calificación o sin ella). • Esta visión coincide parcialmente con otras conceptualizaciones próximas desarrolladas desde la literatura en los últimos años, ante la creciente exploración de las posibilidades prácticas y pedagógicas que ofrece la participación discente en la evaluación:

PARADIGMA TERCERO: EVALUACIÓN COMO INDAGACIÓN

- ✓ Se aproxima al concepto de evaluación compartida sostenido desde la Red de Evaluación Formativa y Compartida, definido como “[...] los procesos dialógicos que mantiene el profesor con su alumnado sobre la evaluación de los aprendizajes y los procesos de enseñanza-aprendizaje que tienen lugar” (A. Pérez, Julián, & López, 2009, p. 37).
- ✓ Dependiendo de la forma que adopte, también puede llegar a incluir la negociación curricular, que tiene por objeto implicar al alumnado en decisiones organizativas de la asignatura, tradicionalmente reservadas al profesorado, incluidas aquellas que afectan a la evaluación (V. M. López, 2011; V. M. López et al., 2007).
- ✓ También se encuentra estrechamente relacionada con la *evaluación democrática* (Batalloso, 2004). Insistiendo en las implicaciones éticas de la evaluación, en este caso, a la participación discente en la evaluación se le suman otras condiciones elementales como la transparencia y los principios de información y explicitación, el diálogo, la negociación, la justicia y la asunción compartida del poder, las decisiones y la responsabilidad (V. M. López, 2006; A. Pérez et al., 2009).

Evaluación auténtica

- Establece que la evaluación debe basarse en la demostración activa del aprendizaje. En consecuencia, la recogida de evidencias se realiza por medio de tareas y situaciones lo más próximas posible a la realidad (Ahumada, 2002; Biggs, 2005; Fook & Sidhu, 2010; Mateo, 2000, 2008; Monereo, 2009; Wiggins, 1989, 1996).

“[...] Es aquella en la que los alumnos son evaluados en función de su actuación activa al utilizar sus conocimientos de manera creativa para resolver problemas plausibles. Tienen que ser problemas reales, es decir, problemas complicados, no rutinarios, y problemas que puedan encontrarse en el campo de estudio o en la vida real de adultos”. (Nevo, 1997, p. 99).
- Esta imagen de la evaluación a menudo aparece sujeta a otras denominaciones como evaluación directa, de la ejecución o alternativa (AR@HKU, 2009b; Mueller, 2004; Nevo, 1997). Esta última denominación se encuentra ampliamente extendida y responde al origen de esta concepción evaluativa como *alternativa* o reacción a la preeminencia de los test o pruebas objetivas (Herman, Aschbacher, & Winters, 1992).
- No obstante, el presente trabajo opta por la evaluación auténtica como denominación de referencia y reserva la evaluación alternativa para designar un enfoque más amplio presentado en el próximo apartado. Asimismo, en línea con Meyer (1992b), Torrance (1995) o Durán (2009), se emplean de forma diferenciada los términos evaluación de la ejecución y la evaluación auténtica.

Evaluación de la ejecución, actuación o desempeño

- Estrechamente relacionada con la evaluación auténtica, la evaluación del desempeño, la ejecución o la actuación también pretende valorar el aprendizaje en acción. Se diferencia de la conceptualización anterior en el grado de realismo de las situaciones que, en vez de auténticas, son verosímiles (podrían existir dada su proximidad con la realidad práctica) (Durán, 2009; Meyer, 1992b).

La evaluación basada en ejecuciones y la auténtica son dos modalidades dentro del denominado enfoque alternativo, en el cual la evaluación de ejecución requiere que los estudiantes, en un momento determinado, demuestren, construyan, desarrollen un producto o solución a partir de unas condiciones definidas y estándares, y la auténtica se basa en la realización de tareas reales. (Mateo, 2008, p. 15).

PARADIGMA TERCERO: EVALUACIÓN COMO INDAGACIÓN

Evaluación basada en competencias

- Frente al modelo formativo tradicional centrado en la enseñanza de contenidos, la introducción de las competencias en el currículum conlleva cambios en las prácticas evaluativas y favorece el desarrollo de nuevas conceptualizaciones.
- Las competencias pasan a formar parte de los planes de estudio dando lugar a modelos formativos *de* competencias, centrados en la acreditación de las mismas, y *por* competencias, más orientados a su desarrollo (A. Díaz, 2006; Villardón, 2006). En respuesta a las nuevas necesidades que plantean, comienzan a emerger distintos modelos para su evaluación: modelos piramidales basados en niveles de desempeño (Miller, 2001) o resultados (Voorhes, 2001) y modelos holísticos o integrados basados en evidencias (Hager & Gonzi, 1994).
- Por encima de las diferencias, las propuestas de evaluación basadas en competencias comparten dos rasgos en todo caso (M. J. García, 2010): la referencia constante a una serie de criterios definidos de antemano y el desarrollo de prácticas situadas, que provoquen la movilización de recursos de naturaleza diversa (conceptuales, procedimentales, y actitudinales). Este último rasgo provoca la conexión de esta conceptualización con la evaluación auténtica y del desempeño.

Evaluación sostenible u orientada al aprendizaje

- Parte de las contribuciones de Boud (2000), que llama la atención sobre el papel que juega la evaluación en el desarrollo de la capacidad de aprendizaje permanente en el alumnado. También se denomina evaluación orientada al aprendizaje (Carless, 2007; Carless, Joughin, & Mok, 2006), entendiendo que dicha capacidad es permanente y que debe ejercerse a lo largo de toda la vida. Clark (2012) habla de evaluación basada en el aprendizaje en el mismo sentido.
- A pesar de su proximidad conceptual, existen ciertos matices que diferencian la evaluación para el aprendizaje y la evaluación orientada al aprendizaje (V. M. López, 2012). Ambas conciben la evaluación como herramienta para la promoción del aprendizaje, pero la evaluación orientada al aprendizaje proyecta su efecto a largo plazo. Este matiz lleva a Boud y Falchikov (2006) a establecer la diferencia entre la evaluación como aprendizaje y como impulso para el aprendizaje para toda la vida. “La evaluación sostenible implica compatibilizar los resultados de aprendizaje actuales con la adquisición de estrategias para aprender de forma autónoma en el futuro” (Gómez, Rodríguez, & Ibarra, 2013, p. 199). Tiene vocación emancipatoria (Buscà, Capllonch, Cladellas, Calvo, Martín, & Padrós, 2010).
- En consecuencia, insiste particularmente en la participación del alumnado en la evaluación, de modo que desarrolle la capacidad de autorregulación y gestión necesaria para seguir aprendiendo a lo largo de la vida. Esta capacidad no es innata y la evaluación debe representar un recurso formativo para su desarrollo, mediante experiencias estructuradas en las que se someta a valoración el propio aprendizaje o el de compañeras/os (I. Álvarez & Olivera-Smith, 2014; Boud, 2000; A. Brew, 2003; Sanmartí, 2010; Sluijmans, Dochy, & Moerkerke, 1999). Ello provoca un enlace directo entre esta conceptualización y la evaluación autónoma o participativa.
- Sin perjuicio de lo anterior, en los últimos años, han aumentado las referencias a la evaluación sostenible desde una perspectiva más práctica, enraizada en su valor instrumental. Así se constata en Liu y Carless, (2006), Carless, Salter, Yang y Lam (2011), Hounsell (2007) o Carless (2013a), que encuentran en la participación discente un apoyo a la viabilidad operativa de los sistemas de evaluación para el aprendizaje.

La presente investigación concibe la evaluación como una herramienta de indagación, entendiendo que la comprensión que genera puede ser empleada como recurso didáctico para la promoción del aprendizaje. Adopta la evaluación para el aprendizaje como

paraguas conceptual e integra el resto de conceptualizaciones como alternativas para favorecer su desarrollo. En consecuencia, la evaluación basada en competencias, la participativa, la sostenible, la auténtica o la de la ejecución representan en este trabajo vías complementarias desde las que provocar la estimulación deliberada del aprendizaje, no sólo su medición o acreditación.

Todas estas conceptualizaciones, forman parte del enfoque evaluativo alternativo, denominado así por su distancia respecto del tradicional, más apegado a los tres paradigmas restantes y a las dos primeras generaciones (ver figura 9). Dichos enfoques, también son producto de la evolución histórica del fenómeno evaluativo y serán objeto de análisis en el próximo apartado.

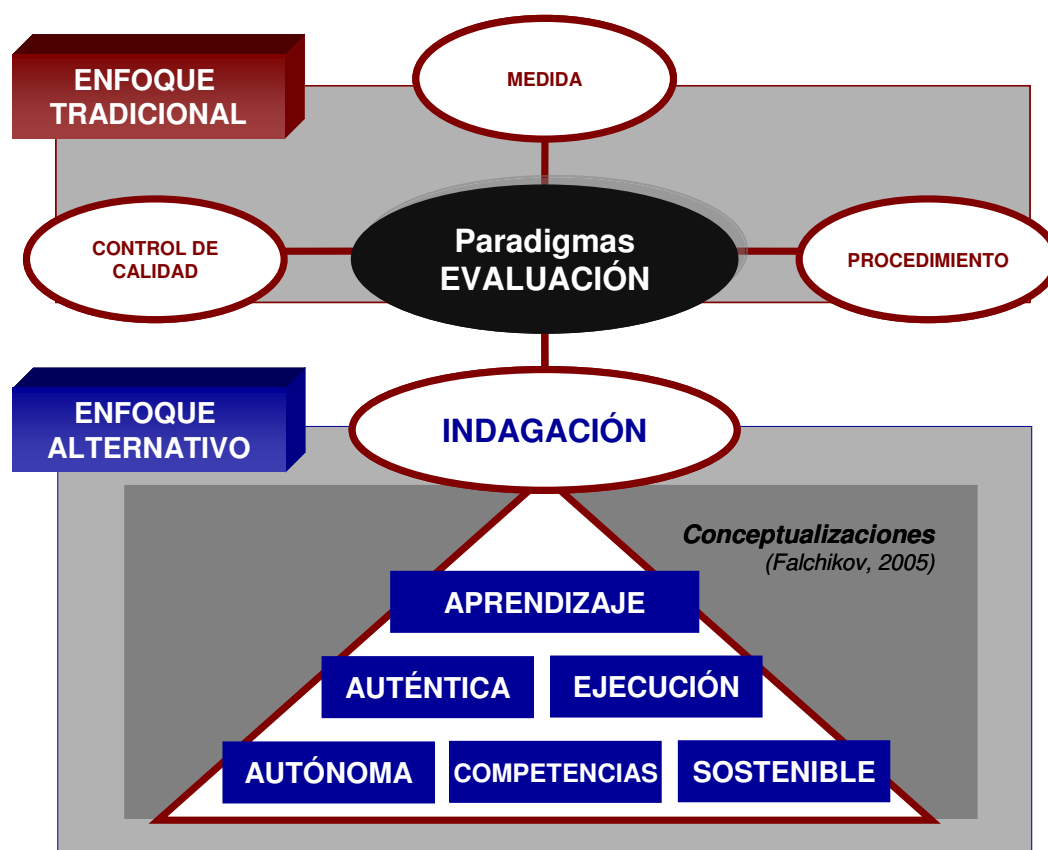


Figura 9. Paradigmas de la evaluación, conceptualizaciones asociadas y principales enfoques.

1.2 Enfoques de evaluación.

El recorrido de las cuatro generaciones descritas ha dado lugar a propuestas evaluativas diversas, propias de cada momento. Con frecuencia se hace referencia al

surgimiento de distintos modelos de evaluación (House, 1980; Monedero, 1998; Monereo & Castelló, 2009b; Stufflebeam & Shinkfield, 1987). Sin embargo, considerando que muchas de las propuestas no alcanzan el rigor y el nivel de desarrollo que formalmente corresponde a los modelos, hay autores que prefieren hablar de creencias (Nevo, 1997) o de enfoques modélicos de evaluación (T. Escudero, 2003).

Calificadas de una u otra manera, la evidente profusión de propuestas evaluativas (en especial, a partir de la tercera generación) ha llevado a muchas/os autoras/es a tratar de agruparlas y sintetizarlas en función de los rasgos que comparten y de sus principales discrepancias. Dichas categorizaciones se conciben como enfoques de evaluación en el presente trabajo, entendiendo que representan grandes tendencias en la forma de emplear y llevar a cabo el proceso evaluativo, basadas en una serie de supuestos de partida. En palabras de Coll, Barberá y Onrubia (2000), representan “el equivalente del pensamiento pedagógico del profesor en lo que concierne a las prácticas de evaluación” (p. 24). Como anticipaba la figura 8, estos enfoques se apoyan en los paradigmas de evaluación, que son las creencias que los motivan, y se manifiestan en el transcurso de las cuatro generaciones evaluativas.

Entendidos desde esta perspectiva, existe un cierto acuerdo al identificar dos grandes enfoques en la historia de la evaluación del aprendizaje: el clásico o tradicional y, por oposición, el caracterizado como alternativo (Falchikov, 2005; Wolf & Reardon, 1996). La tabla 8 da cuenta de los principales rasgos de cada uno de ellos, siendo el segundo el elegido como referencia en la presente investigación.

Tabla 8

Comparación de los enfoques evaluativos tradicional y alternativo. Revisión ampliada de la propuesta de Wolf y Reader (1996)

PRINCIPALES ENFOQUES DE EVALUACIÓN	
TRADICIONAL	ALTERNATIVA
- Parte de la lógica positivista del aprendizaje y su evaluación. Incluye propuestas evaluativas desarrolladas desde los paradigmas de la medida, el procedimiento y el control de calidad, de modo que se corresponde con las primeras generaciones del desarrollo evolutivo del concepto. - Incluye propuestas evaluativas dirigidas a medir un aprendizaje que se supone objetivable. Su desarrollo práctico se caracteriza por el empleo de técnicas evaluativas cuantitativas, en las que la respuesta y la valoración es única y unívoca (test de opción múltiple, de verdadero o falso,	- Surge como reacción al traslado de la lógica positivista al aprendizaje y su evaluación. Alberga las propuestas surgidas desde el paradigma evaluativo como indagación e investigación. Su desarrollo tiene lugar en las dos últimas generaciones, la del juicio y la respondente-constructivista. - Incluye propuestas evaluativas dirigidas a comprender el aprendizaje, partiendo de su complejidad y de la imposibilidad de devolver una valoración plenamente objetiva. Más que la medición, persiguen la construcción de un juicio profesional por medio de técnicas variadas, principalmente cualitativas

PRINCIPALES ENFOQUES DE EVALUACIÓN	
TRADICIONAL	ALTERNATIVA
pruebas de emparejamiento, etc.). • Incluye propuestas evaluativas que focalizan la atención en puntuaciones estandarizadas que permiten establecer comparaciones normativas. • Unido a lo anterior, la evaluación produce información que puede generalizarse. • La evaluación queda en manos de agentes externos o del profesorado. • La evaluación se plantea como un anexo final que desempeña una función puntual. • Incluye propuestas evaluativas lineales (Tejedor, 1999), en las que la información evaluativa no se emplea como recurso para la adopción de decisiones de mejora docente, institucional o curricular. No hay reacción o actualización. • La evaluación aporta información fundamentalmente cuantitativa que atiende sólo a la comunicación del resultado (puntuación).	(observación, resolución de problemas, proyectos, etc.). • Las propuestas incluidas focalizan la evaluación en el aprendizaje del alumnado de manera individualizada, sin referencia a normas. • Unido a lo anterior, la evaluación se considera un conocimiento idiosincrásico. • La evaluación se desarrolla contando con la participación activa del alumnado. • La evaluación está integrada en el proceso formativo, no es un apéndice. • Incluye propuestas evaluativas circulares, interactivas o recursivas (Bordas & Cabrera, 2001; Tejedor, 1999), en las que la información evaluativa facilita la acción curricular y la adopción de medidas de corrección y mejora. • La evaluación aporta información relevante sobre el proceso formativo; las conclusiones alcanzadas se acompañan de comentarios que facilitan su comprensión y de sugerencias de mejora para próximos desempeños.

La evaluación tradicional representa un término general vinculado a procedimientos convencionales de evaluación (como test, pruebas de opción múltiple o ensayos), que se caracterizan por primar el aprendizaje pasivo y el recuerdo y repetición de la información (AR@HKU, 2009b). A. Pérez et al. (2009) definen la evaluación alternativa como aquella que “hace referencia a todas las técnicas y métodos de evaluación que intentan superar la metodología tradicional de evaluación, basada en la simple realización de pruebas y exámenes, con la finalidad única o principal de calificar” (p. 32). En consecuencia, todo procedimiento evaluativo que se desmarque del uso de los recursos clásicos puede considerarse propiamente como fórmula alternativa (Mateo, 2000). Esta es la causa de que conceptualizaciones diferentes y más precisas como la evaluación participativa o autónoma, la auténtica o de la ejecución sean frecuentemente designadas como evaluación alternativa en la literatura (Ahumada, 2002; I. Álvarez, 2008b; Herman et al., 1992; Knight, 2005; Nevo, 1997).

2. Construcción de una definición consensuada: ¿qué entendemos por evaluación?

La divergencia de interpretaciones representa una dificultad clásica en la literatura y la investigación evaluativa. En consecuencia, varios autores han acometido la tarea de analizar las numerosas definiciones existentes, para detectar elementos clave en la construcción de un significado compartido:

- Castillo y Cabrerizo (2003) revisaron una muestra de 23 definiciones alcanzando las siguientes conclusiones con respecto a la evaluación:
 - ✓ Representa un componente fundamental del proceso educativo.
 - ✓ Se trata de un proceso complejo, abierto y contextualizado, no de una mera acción puntual.
 - ✓ Puede servir a propósitos diversos partiendo de la obtención y el análisis informativo: formular juicios de valor, tomar decisiones y retroalimentar.
- Analizadas varias propuestas, Alegre y Villar (2006) señalaron cuatro claves para clarificar el significado de la evaluación: la recopilación sistemática de información, la variedad de objetos evaluables, la diversidad de propósitos y el empleo efectivo de las conclusiones obtenidas. La definición de Patton (1982) integra todas ellas con sentido:

La práctica de la evaluación implica la colección sistemática de información acerca de las actividades, características y resultados de programas, de personal y de productos usados por la gente con objeto de reducir las incertidumbres, mejorar la efectividad, y tomar decisiones respecto de lo que hacen y cómo afectan los programas, el personal, o los productos. (p. 15).
- R. López (2001) también realizó un intento de síntesis en el desarrollo de su investigación doctoral, concluyendo que en las distintas conceptualizaciones se repiten tres elementos comunes:
 - ✓ La evaluación se refiere a una capacidad humana.
 - ✓ Culmina en una toma de decisiones. Tiene carácter instrumental, no es un fin en sí misma.

- ✓ Implica un juicio de valor sobre el objeto evaluado y éste debe reposar sobre unos criterios claramente establecidos que eviten la improvisación.
- Monescillo y Toscano (2007) llevaron a cabo un análisis de las definiciones ofrecidas al término evaluación desde generaciones y paradigmas diferentes. A modo de conclusión, presentaron su propia definición de la evaluación como:

[...] proceso necesario, intencional, continuo, planificado, sistemático y de finalidad múltiple, que aporta informaciones de todos y cada uno de los elementos implicados en el proceso educativo, que ayudan a comprender el desarrollo de los procesos formativos y facilitan la toma de decisiones curriculares. (p. 57).

Para favorecer la univocidad de interpretaciones en sus propósitos y conclusiones, el presente trabajo parte de la delimitación de una propuesta propia de conceptualización del fenómeno evaluativo. Su configuración surge de la confluencia de dos referencias complementarias:

- Por un lado, se nutre de la perspectiva general que proporcionan los trabajos de síntesis reseñados, asumidos como muestra de los acuerdos elementales sobre el tema.
- Por el otro, parte de la posición adoptada con respecto al fenómeno evaluativo en el primer apartado. En consecuencia, plantea una definición propia de la cuarta generación y próxima al enfoque alternativo, donde la evaluación representa un medio de indagación al servicio del aprendizaje.

Entendemos por evaluación el...

proceso sistemático de recogida y análisis de información y evidencias sobre una serie de criterios delimitados de antemano, para comprender un objeto, fenómeno o situación determinada y actuar en consecuencia.

La definición propuesta está formulada desde una perspectiva funcional e integra siete elementos clave que caracterizan el significado conferido a la evaluación. Como recoge posteriormente la tabla 9, estas claves también se encuentran presentes en otras definiciones reseñadas en la literatura:

● ***La evaluación es un proceso sistemático.***

Una primera conclusión que se desprende del análisis de las definiciones formuladas desde diferentes generaciones y paradigmas es que la evaluación es un proceso, no un acto puntual y aislado, ni una actividad simple y mecánica. Este rasgo representa un acuerdo general en la literatura y está presente desde las primeras contribuciones a la consolidación de la evaluación educativa, las asociadas al trabajo de Tyler (1950).

Como será examinado en el próximo apartado, entender la evaluación como proceso implica reconocer que es constitutiva de fases o momentos diferentes.

● ***La motivación inicial de toda evaluación es la comprensión.***

Adoptado el paradigma de la indagación como visión evaluativa de referencia, entendemos que la motivación inicial de toda evaluación es la comprensión de aquello que se evalúa. Sin embargo, dicha comprensión no representa un fin en sí misma, y se pone al servicio de una amplia gama de finalidades prácticas que enlazan con la actuación.

● ***La comprensión que genera la evaluación se emplea para actuar en relación a finalidades específicas más concretas.***

Junto a la motivación inicial de la comprensión, la conceptualización propuesta incorpora la actuación a partir de la misma como elemento clave. En las definiciones y síntesis consultadas, esta idea se concreta en la toma de decisión que sucede a la evaluación. Así se constata tanto en definiciones propias de la tercera generación (Coll, 1983; Cronbach, 1963; Patton, 1982; Stufflebeam et al., 1971; Tenbrink, 1981), como en aquellas otras más recientes sujetas a la cuarta (Mateo, 2000; Sanmartí, 2007; Yániz & Villardón, 2006).

Esta tercera clave hace referencia al carácter instrumental de la evaluación que recoge de forma expresa R. Fuentes (2001). Lejos de constituir un fin en sí misma, la evaluación debe servir a unos propósitos determinados (Ahumada, 2002; A. Pérez et al., 2009; Rizo, 2004). La comprensión que proporciona puede ponerse al servicio de decisiones o propósitos diversos, sociales o pedagógicos, pero debe ser empleada en cualquier caso (Alegre & Villar, 2006).

● ***La evaluación exige llevar a cabo una recopilación consciente de evidencias.***

Concebida como investigación, la evaluación ha de servirse de la recogida de toda una serie de evidencias (datos, muestras, informaciones, etc.) que le ayuden a la

consecución de su primera motivación, la comprensión de su objeto de estudio. Las evidencias son pruebas determinantes (Real Academia Española, 2014) y su naturaleza puede ser muy diversa en función del objeto evaluativo concreto al que hagan referencia. Su recopilación debe ser cuidadosamente planificada y practicada con rigor, como corresponde a la sistematicidad del proceso evaluativo.

● ***La evaluación exige llevar a cabo un análisis consciente de las evidencias recogidas.***

La recopilación de evidencias no garantiza por sí misma la comprensión de aquello que pretende evaluarse, de modo que es necesario llevar a cabo una tarea de análisis e interpretación a partir de las mismas. Así lo exige concebir la evaluación como proceso sistemático de investigación.

● ***La evaluación reposa sobre una serie de criterios definidos de antemano.***

La recogida de evidencias y su posterior análisis se lleva a cabo en referencia a una serie de criterios que delimitan los aspectos concretos a valorar. Como proceso sistemático, la evaluación debe alejarse de la dispersión e imprecisión en su empeño por facilitar la comprensión del objeto al que se aproxima. Para ello, debe partir del establecimiento de una serie de referencias que orienten la recopilación de evidencias y su posterior análisis, y así se reconoce desde definiciones muy dispares (Ahumada, 2005; De la Orden, 1985b; R. Fuentes, 2001; Tyler, 1950; Yániz & Villardón, 2006).

Los criterios de evaluación representan la referencia comparativa que permite comprender la calidad del objeto evaluado y adoptar las decisiones oportunas (Camilloni, 1998; Nevo, 1997). Expresan las características que se esperan y que es preciso valorar en un determinado proceso, producto o servicio (Tejedor, 2003). A menudo se acompañan de indicadores de evaluación, que recogen aspectos representativos, observables y verificables en los mismos. Estos últimos proponen datos empíricos de naturaleza cualitativa o cuantitativa que permiten estimar la productividad y funcionalidad del sistema (De Miguel, 2000).

Dependiendo del marco en el que se lleve a cabo la evaluación, dichos criterios pueden ser complementados o reemplazados por principios de calidad comúnmente acordados por la comunidad profesional del ámbito de referencia, esto es, por estándares (JCSEE, 1998, 2001, 2007). Estos últimos también responden a la necesidad de centrar la recogida de evidencias y proporcionar criterios comparativos para su análisis y valoración, pero aportan mayor rigor al proceso evaluativo por ser

fruto de procesos de negociación y consenso internacional (Baeza, Pérez, & Reyes, 2006).

● **La evaluación puede atender a objetos, fenómenos o situaciones diversas.**

La última de las claves presente en la definición propuesta, refiere a la variedad de focos de atención a los que puede aproximarse el proceso evaluativo. Interpretada como indagación o investigación, la evaluación gira en torno a un objeto de estudio, aquel aspecto de la realidad que desea entender para luego actuar en consecuencia. Habitualmente se denomina objeto de evaluación o evaluando (Scriven, 1967; Stake, 2006).

En el contexto educativo, los objetos de evaluación han estado tradicionalmente apegados al alumnado y al profesorado. Como indica Nevo (1997), la evolución del significado atribuido a la evaluación ha dado lugar a la emergencia progresiva de nuevos focos de atención, orientados a la evaluación de materiales didácticos, programas, proyectos o centros. Así lo confirman las aportaciones realizadas por el JCSEE a la definición de estándares para la evaluación de proyectos y materiales, personal, programas o aprendizajes (JCSEE, 1981, 1998, 2001, 2007).

La presente investigación se ciñe al más clásico de los objetos de evaluación referidos: el aprendizaje discente. Sin negar la relevancia de otros focos y espacios de atención, estudia la evaluación que lleva a cabo el profesorado universitario del aprendizaje del alumnado en un espacio formativo en particular (los grados para la formación básica de maestras/os de Educación Infantil y Primaria). La adopción de este centro de interés conlleva una serie de implicaciones específicas que se resumen a continuación:

- ✓ En los países de habla inglesa se distingue entre los términos *evaluation* (evaluación referida a centros, programas, materiales, etc.) y *assessment* (evaluación del aprendizaje) (Brockbank & McGill, 2002; Harlen, 2007; Moleenar et al., 2009). La acotación planteada nos sitúa ante esta última denominación.
- ✓ En el ámbito universitario la evaluación del aprendizaje del alumnado se ha reducido a la evaluación del rendimiento académico, hasta el punto de llegar a considerarlas la misma cosa (Jonsson, 2013; Smitter & Siso, 2007). Se trata de una generalización muy extendida. Prueba de ello es que el *Diccionario de la Lengua Española* (Real Academia Española, 2014) define la

evaluación como la estimación de los conocimientos, las aptitudes y el rendimiento del alumnado en una de sus acepciones.

- ✓ El aprendizaje representa un proceso mucho más amplio y complejo que lo que deja ver una calificación. En consecuencia, son frecuentes las críticas que señalan las limitaciones de su equiparación con el rendimiento (T. Escudero, 1987; S. Rodríguez, 1982). Entre otros efectos, reduce el aprendizaje a resultados individuales e inmediatos, desconsiderando otro tipo de logros (De la Orden, 1985b) y lo simplifica, hasta el punto de considerarlo mensurable en una evaluación cuantitativa y final (G. Pérez, 1986). Como concluye Pérez Juste¹⁵ (1991), esa asimilación responde sólo a circunstancias prácticas y operativas, sujetas a las posibilidades de la investigación.
- ✓ Incluso representando una reducción operativa, la evaluación del rendimiento es tan amplia y variada como los objetivos educativos a los que puede responder (Tejedor, 1999). No existe una única manera de interpretar el rendimiento académico, de modo que en la investigación específica se evidencian definiciones y orientaciones dispares (Pérez Juste, 1991). Mientras en determinados estudios el rendimiento se interpreta como resultado o producto¹⁶, en otros se enfoca como nivel¹⁷.
- ✓ Frente a este abanico de interpretaciones y limitaciones, la presente investigación plantea la evaluación del aprendizaje como la evaluación del rendimiento académico discente, pero sólo en la medida en la que atiende a su manifestación externa, y no a los procesos mentales internos que median

¹⁵ La referencia se realiza utilizando los dos apellidos del autor para poder diferenciarlo de Rafael Pérez Pérez.

¹⁶ Es la definición operativa más recurrida y empleada en la investigación evaluativa y así lo constatan Valentine, DuBois, & Cooper (2004) en un trabajo de metanálisis realizado sobre 60 publicaciones que relacionan el rendimiento académico con creencias de autoeficacia del alumnado. A pesar de las críticas, esta definición está presente en investigaciones desarrolladas décadas atrás (Gimeno 1976; cit. en Tejedor, 1999; J. C. Núñez et al., 1998) y en otras más recientes (Vandevoort, Amrein-Beardsley, & Berliner, 2004; De la Fuente, Vicente, Sánchez, & Berbén, 2010; De la Fuente, Sander, & Putwain, 2013; A. González, Donolo, Rinaudo, & Paoloni, 2011).

¹⁷ A ella responden los estudios que recurren al cálculo de tasas de rendimiento en las que se consideran de forma ponderada una serie de factores vinculados (éxito, fracaso, relación de créditos matriculados y superados, número de intentos, etc.). Este recurso, lo hallamos en Martín, García, Torbay y Rodríguez (2008) o en Pérez Juste (1991). También se usa de forma generalizada para analizar y comparar los sistemas educativos.

en su consecución (Rosales, 1988; cit. en Tejedor et al., 1999). En nuestro trabajo el rendimiento no se limita a tasas o puntuaciones concretas, sino que hace referencia al desarrollo progresivo que realiza el alumnado de las competencias previstas, según se pone de manifiesto en las actividades evaluativas acreditativas y pedagógicas desarrolladas a lo largo del proceso formativo. Esta ampliación de sentido es posible porque el aprendizaje (o rendimiento académico) no representa una variable dependiente en el estudio, de modo que no es preciso proporcionar sobre la misma una medición comparable. Su interés es indirecto y cualitativo, ya que el núcleo central de la investigación es la competencia evaluativa del profesorado, no el aprendizaje sobre el que se aplica.

En síntesis, la propuesta expuesta sólo representa una definición operativa y de mínimos, ideada para centrar el trabajo y posibilitar la transferencia de sus conclusiones. Recoge para ello las ideas en las que mayor acuerdo revelan las síntesis de definiciones consultadas. De acuerdo con la posición específica adoptada con respecto a los grandes paradigmas de la evaluación, también incorpora de forma selectiva los aspectos referidos en conceptualizaciones de generaciones distintas. La tabla 9 muestra de forma sintética la representación de los siete elementos clave de nuestra definición en ambos tipos de referencia.

Tabla 9
Representación de las claves de la definición evaluativa adoptada en trabajos previos de síntesis y otras conceptualizaciones

TRABAJOS DE SÍNTESIS	IDEAS CLAVE CONCEPTUALIZACIÓN						
	1	2	3	4	5	6	7
Castillo & Cabrerizo (2003)	X	X	X	X	X		
Alegre & Villar (2006)	X	X	X	X			X
R. López (2001)	X	X	X			X	X
Monescillo & Toscano (2007)	X	X	X				
OTRAS DEFINICIONES	1	2	3	4	5	6	7
Tyler (1950)						X	
Cronbach (1963)			X	X			
Scriven (1967)		X					X
Stufflebeam et al. (1971)			X	X			
Tenbrik (1981)		X	X	X			
JCSEE (1981, 2007)		X					X

	IDEAS CLAVE CONCEPTUALIZACIÓN						
Patton (1982)	X	X	X	X			X
Coll (1983)		X	X			X	X
De la Orden (1985a)		X	X	X	X		X
Stufflebeam & Shinkfield (1987)		X					X
Casanova (1992)	X	X	X	X			
Nevo (1997)		X		X			X
Mateo (2000)	X		X	X			
Pérez Juste (2002)	X	X	X	X	X	X	X
Ahumada (2005)		X	X	X	X		
Sanmartí (2007)		X	X	X			
Yániz & Villardón (2006)	X	X	X	X	X	X	X

Nota: Las claves comprobadas se representan en el mismo orden que han sido expuestas: (1) Proceso sistemático; (2) Orientada a la comprensión (3) Comprometida con la actuación; (4) Recogida de evidencias; (5) Análisis de evidencias; (6) Basada en criterios; (7) Aplicable a objetos variados.

Las coincidencias señaladas responden a un criterio de literalidad. El análisis efectuado, sugiere que la ausencia de ciertas claves responde más a la plena asunción de las mismas, que su cuestionamiento.

El resumen realizado sugiere que la definición adoptada ofrece una buena síntesis de las diversas visiones sostenidas sobre la evaluación, en especial, de aquellas más recientes formuladas en la cuarta generación evaluativa. De todas las consideradas, la visión que más se aproxima a la definición configurada es la que aportan Yániz y Villardón (2006), que conciben la evaluación como:

Conjunto de actividades que conforman un proceso sistemático de recogida, análisis e interpretación de información, que conduce a emitir un juicio sobre las realizaciones de una persona, grupo, objeto, situación o fenómeno a partir de la comparación con unos criterios previamente establecidos y con vistas a tomar una decisión. (p. 78).

Pérez Juste (2002) también ofrece una visión de la evaluación plenamente compatible, entendida como “valoración, a partir de criterios y referencias preespecificados, de la información técnicamente diseñada y sistemáticamente recogida y organizada, sobre cuantos factores relevantes integran los procesos educativos para facilitar la toma de decisiones de mejora” (p. 57).

3. Proceso de evaluación del aprendizaje

Según la definición presentada, la evaluación es un proceso sistemático en el que se entrelazan un número variable de fases interdependientes. Interpretado desde el paradigma de la indagación, dicho proceso se asemeja de manera evidente al que tiene lugar en cualquier investigación, aunque con ciertas salvedades. Mientras que la investigación se orienta a la obtención de conclusiones más o menos generalizables, la finalidad de la evaluación es comprender una realidad particular para poder actuar en consecuencia (Casanova, 1992).

Al examinar la secuencia de actuación que se propone para llevar a cabo la evaluación del aprendizaje discente, la literatura da cuenta de dos tendencias paralelas: la delimitación general de grandes fases y/o el análisis pormenorizado de los pasos que conlleva su desarrollo. En una fusión de ambas, las próximas páginas recorren las principales etapas del proceso y la secuencia interna de tareas y pasos previstos en cada una de ellas.

La mayor parte de las propuestas consultadas identifican tres grandes momentos asociados a la evaluación: el diseño o planificación, la implementación o puesta en práctica y el establecimiento de conclusiones. La presente investigación añade a ellos la metaevaluación o revisión de la evaluación, una cuarta etapa que no se encuentra tan extendida ni generalizada:

● **Planificación o diseño de la evaluación.**

Se trata de la fase de delimitación inicial que debe tener lugar para garantizar la sistematicidad del proceso evaluativo. Entraña toda una serie de elecciones y decisiones sobre el planteamiento de la evaluación y su desarrollo, que deben adoptarse conscientemente, atendiendo a los propósitos que se persiguen y a las dificultades previsibles en su consecución. El diseño o planificación de la evaluación constituye una etapa clave del proceso evaluativo y su relevancia suscita un acuerdo unánime en la literatura específica.

Integrando las aportaciones iniciales de varios autores, entendemos que esta primera fase implica recorrer los once pasos o tareas recogidas en la tabla 10. La división entre las mismas se representa con líneas discontinuas para reflejar su interdependencia.

Tabla 10
Secuencia interna de pasos en la planificación evaluativa

DISEÑO O PLANIFICACIÓN DE LA EVALUACIÓN

Delimitar los objetivos de la evaluación (función, propósitos específicos).

Aclarar los propósitos específicos que persigue la evaluación o la función a la que pretende servir la comprensión generada (social o pedagógica, según se expone en el próximo apartado).

Definir el objeto de evaluación: criterios, indicadores o en su caso, estándares.

Determinar con la mayor precisión posible los aspectos concretos que pretenden evaluarse y sobre los que debe tener lugar la recogida de evidencias y su análisis posterior. Tal delimitación puede llevarse a cabo por medio de un sistema de criterios e indicadores de evaluación o, en su caso, estándares.

Seleccionar evidencias adecuadas y suficientes para probar la consecución de los criterios e indicadores establecidos.

De acuerdo con el sistema de criterios e indicadores formulado, prever un repertorio adecuado y suficiente de pruebas o evidencias (datos e informaciones, producciones de distinto tipo, demostraciones, etc.) que puedan servir para su corroboración.

Seleccionar las tareas e instrumentos de evaluación que se propondrán para recopilar evidencias necesarias.

Seleccionar (diseñar, adecuar o elegir entre recursos existentes) tareas e instrumentos de evaluación para recopilar las evidencias previstas. De acuerdo a lo precisado en el primer paso, las situaciones o tareas evaluativas pueden tener carácter pedagógico y/o acreditativo según se estime en cada caso.

Determinar momentos precisos para llevar a cabo la recogida de evidencias y designar los agentes evaluados y evaluadores en cada caso.

Delimitar los momentos precisos en los que se llevarán a cabo las tareas previstas (recogida de evidencias) y la distribución de responsabilidades en torno a las mismas (agente evaluador y evaluando y condiciones generales de la participación en cada caso).

Prever los análisis que se llevarán a cabo sobre las evidencias recogidas y las herramientas o recursos necesarios para su desarrollo.

Anticipar el procesamiento y análisis que será llevado a cabo sobre las evidencias una vez se hayan recopilado, configurando las herramientas o recursos necesarios (bases de datos para tratamiento estadístico, plantillas de corrección, rúbricas, etc., según caso). Dichas herramientas facilitan la tarea de análisis y evitan desviaciones en su desarrollo.

Prever las dificultades y complicaciones previsibles y posibles recursos para hacerles frente.

Identificar las circunstancias prácticas que puedan obstaculizar el desarrollo del plan de evaluación previsto y la consecución de sus objetivos en condiciones razonables. Anticipar posibles alternativas para contrarrestar su impacto.

Prever mecanismos y medidas para favorecer la difusión de las condiciones de la evaluación entre el alumnado y la comprensión de los criterios e indicadores.

Planificar el modo de presentar a las/os estudiantes los criterios e indicadores a los que deben aproximar su trabajo y las condiciones que regirán su evaluación (previsión de momentos, preparación de documentación y posibles modelos, herramientas, etc.). Anticipar medidas para favorecer su comprensión e interiorización, no sólo la información.

Prever procedimientos y vías para llevar a cabo la comunicación de los resultados y procesos de feedback o retroalimentación a partir de los mismos.

Planificar el modo de comunicar las conclusiones de la evaluación (momentos, audiencias interesadas, formatos de comunicación, modelos de informe, etc.) y el trabajo a efectuar a partir de dicha acción informativa. Ello comprende tanto los procesos de retroinformación o feedback (profundizar en la comprensión de los resultados comunicados) como los de proacción (propuestas y sugerencias para la mejora).

Prever mecanismos para favorecer el seguimiento del objeto evaluado a lo largo del proceso (registros y modos de empleo de los mismos).

Decidir el tipo de seguimiento que se pretende realizar del grupo de estudiantes y sus integrantes y crear las herramientas necesarias para su desarrollo (modelos de registro de logros y avances, calendario del seguimiento, etc.).

Planificar la metaevaluación.

Establecer los propósitos de la metaevaluación, definir el objeto de la revisión (criterios, indicadores o estándares), identificar evidencias suficientes y adecuadas para su valoración, seleccionar instrumentos y situaciones para su recopilación, determinar momentos concretos para realizar la recogida y designar agentes evaluados y evaluadores para cada caso.

Nota: El esquema representa una propuesta propia, elaborada revisando e integrando las aportaciones iniciales de De Miguel (2010), Guerrero, Huertas, Mor y Rodríguez (2013), Mateo (2000), Natriello (1987), Tejedor (1999) y Villardón (2006). El sombreado en la última fila indica que la tarea contenida corresponde propiamente a la cuarta fase de metaevaluación, expuesta más adelante.

● **Implementación o puesta en práctica.**

La puesta en práctica o implementación de la evaluación es la consecuencia lógica de la fase de planificación y se concreta en el desarrollo de todas aquellas actividades e iniciativas previstas. Reconocida de común acuerdo, para designarla se emplean denominaciones diferentes. Así, mientras que Mateo (2000) y Villardón (2006) hablan de desarrollo, Tejedor (1999) hace referencia a la ejecución.

Dentro de la puesta en práctica del proceso evaluativo tienen lugar seis grandes actuaciones, interdependientes entre sí y derivadas de las decisiones adoptadas en la fase anterior (ver tabla 11):

Tabla 11
Secuencia interna de pasos en la implementación o puesta en práctica evaluativa

PUESTA EN PRÁCTICA O IMPLEMENTACIÓN DE LA EVALUACIÓN

Verificar el grado ajuste de las condiciones reales de la evaluación y las previsiones iniciales e introducir medidas correctoras, si fuera necesario.

Controlar que la evaluación planificada sea realizable en las condiciones reales e identificar los factores que puedan representar un obstáculo o amenaza en su puesta en práctica (modificación imprevista del calendario, recorte de recursos, riesgos de ilegitimidad no previstos, etc.). Si fuera preciso, esa verificación debería traducirse en el reajuste del plan de evaluación previsto, de forma que no se alteren sus propósitos fundamentales.

Realizar la recogida de evidencias, documentando las posibles incidencias.

Llevar a cabo la recogida de evidencias en las condiciones previstas o reajustadas (herramientas y situaciones, agentes implicados, momentos, etc.) y documentar toda aquella alteración que pueda poner en entredicho su validez o credibilidad como prueba.

Analizar y procesar las evidencias recopiladas.

Practicar el procesamiento y análisis de evidencias previsto, contando con las herramientas y los recursos dispuestos para ello.

Formular los juicios o valoraciones correspondientes, en función de los criterios e indicadores definidos.

Comparar el nivel de logro evidenciado en el análisis realizado, con las referencias previstas en los indicadores formulados en la fase de planificación. Definir el juicio o valoración correspondiente.

Cumplimentar los registros de seguimiento previstos.

Cumplimentar los registros de seguimiento dispuestos en la fase de planificación con las conclusiones y resultados que se vayan generando a nivel grupal y/o individual.

Implementar la metaevaluación.

Valorar las condiciones prácticas reales de la metaevaluación, realizar la recogida de evidencias prevista, efectuar su análisis y formular los correspondientes juicios o valoraciones.

Nota: El sombreado en la última fila indica que la tarea contenida corresponde propiamente a la cuarta fase de metaevaluación, expuesta más adelante.

● ***Establecimiento de conclusiones para la actuación o toma de decisión.***

La tercera fase del proceso implica el establecimiento de las conclusiones de la evaluación de acuerdo a los propósitos para los que haya sido diseñada y desarrollada. La evaluación debe generar comprensión sobre el objeto evaluado (en este caso, el aprendizaje), para poder emplearla al servicio de otros fines, pedagógicos o sociales. Esta es la fase que enlaza con la actuación o decisión, aquella en la que se invierte la información generada.

Como en el caso anterior, representa una etapa evaluativa reconocida de forma unánime, aunque con nomenclatura diferente. Dependiendo del contexto específico se habla de valoración (Tejedor, 1999), contrastación (Mateo, 2000), resultados (Villardón, 2006) o calificación¹⁸ (Monescillo & Toscano, 2007).

En la presente investigación esta fase se concibe sujeta a las cuatro tareas o pasos resumidos en la tabla 12:

¹⁸ Aunque aparezca aquí recogida para representar la variedad de denominaciones, no compartimos esta última. Desde nuestro punto de vista, representa una reducción innecesaria del proceso evaluativo, al desconsiderar las decisiones pedagógicas no acreditativa que en ella puedan tener lugar.

Tabla 12
Secuencia interna de pasos en la planificación evaluativa

CONCLUSIONES DE LA EVALUACIÓN PARA LA ACTUACIÓN O DECISIÓN

Comunicar las conclusiones según lo previsto (reuniones, informes, listados, etc.).

Realizar la comunicación de las conclusiones generadas en la evaluación según las decisiones adoptadas en la fase de planificación (listados, informes personalizados y/o generales, devolución de correcciones, entrevistas o sesiones individuales/grupales, etc.).

Desarrollar los procesos de retroalimentación y proacción pertinentes.

Como será definido más adelante, ello supone desarrollar las iniciativas previstas para profundizar sobre los resultados comunicados (feedback) y establecer sugerencias o pautas para la mejora de los mismos (proacción). Esta última actuación cruza al próximo paso.

Hacer uso de las conclusiones de la evaluación: adoptar las decisiones previstas.

Emprender las acciones o adoptar las decisiones a las que se orientaba la evaluación desde la fase de planificación. Dichas decisiones pueden afectar al aprendizaje discente, a la enseñanza y/o al currículum. De igual modo, pueden representar un desenlace pedagógico (reforzar, ajustar, mejorar, etc.) y/o acreditativo (calificar, promocionar, etc.), según se haya previsto.

Conclusiones de la metaevaluación.

Comunicar y/o documentar las conclusiones, asegurar la retroalimentación y la proacción y adoptar las decisiones pertinentes.

Nota: El sombreado en la última fila indica que la tarea contenida corresponde propiamente a la cuarta fase de metaevaluación, expuesta más adelante.

● **Metaevaluación o revisión de la evaluación.**

Siguiendo a Scriven (1968) entendemos por metaevaluación la evaluación de la evaluación, es decir, su revisión. Esta cuarta etapa representa una nueva evaluación dentro de otra y, tal consideración, conlleva una serie de consecuencias directas:

- ✓ Como cualquier otra evaluación, la metaevaluación debe desarrollarse como proceso, recorriendo las fases de planificación, puesta en práctica y presentación de conclusiones para la actuación o decisión. No es frecuente, pero sobre la misma cabría llevar a cabo una cuarta fase de revisión (una evaluación de la metaevaluación).
- ✓ Como cualquier otro proceso evaluativo, la metaevaluación no puede depender de la improvisación. Debe llevarse a cabo de forma sistemática, delimitando cuidadosamente las decisiones vinculadas a su diseño y puesta en práctica y enlazando las conclusiones con los fines previstos (Stufflebeam, 1974/2011, 2001b).

En consecuencia, la metaevaluación no debe comenzar a plantearse una vez haya finalizado la evaluación cuya revisión pretende. Puede desarrollarse según transcurra el proceso evaluativo o cuando haya concluido (Santos, 2003; Santos & Moreno, 2004), pero con premeditación en cualquier caso. Por ese motivo, las actuaciones correspondientes a esta cuarta fase han sido integradas en las tres anteriores (figuran sombreadas en la tabla 10, tabla 11 y tabla 12), en referencia al diseño, la puesta en práctica y las conclusiones de la revisión.

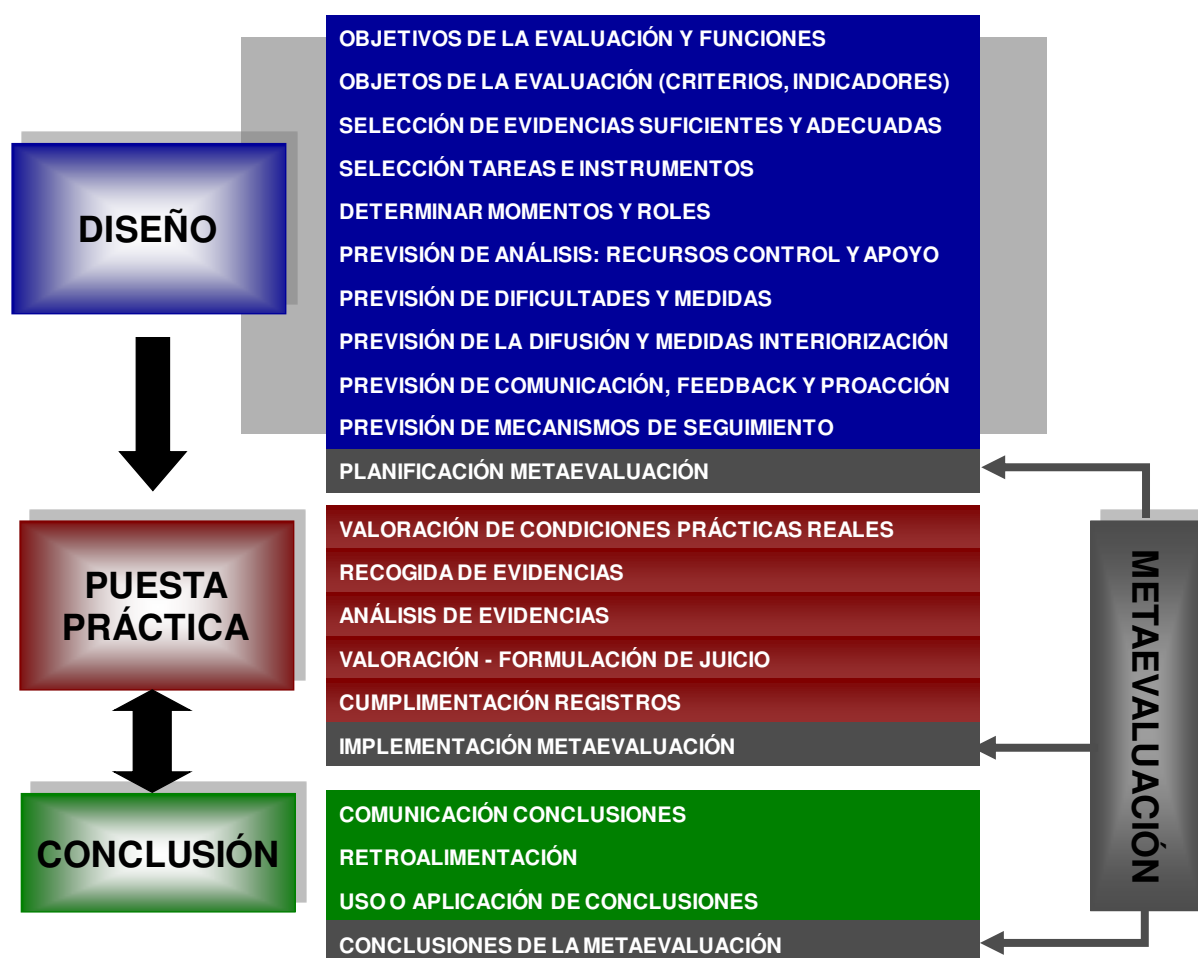


Figura 10. El proceso de evaluación: propuesta flexible de fases y tareas.

El recorrido representado en la figura 10 debe interpretarse desde dos matices relevantes. El primero de ellos hace referencia a la secuencia de las cuatro fases evaluativas, que no se ciñe estrictamente a la linealidad que sugiere su exposición. Admitido que la evaluación no puede representar una práctica puntual y final, las fases segunda y tercera, por ejemplo, se pueden producir de forma repetida y solapada a lo largo de un mismo ciclo formativo. El segundo matiz hace referencia a la flexibilidad del esquema general de fases y tareas enumeradas, que no representa un patrón rígido. Los propósitos

específicos de la evaluación, las condiciones prácticas y operativas o las particularidades del aprendizaje al que se oriente pueden dar lugar a adecuaciones en el recorrido descrito.

4. Funciones de la evaluación del aprendizaje

La comprensión representa la motivación inicial compartida en toda evaluación que se desarrolle desde la perspectiva de la indagación. “Es decir, se evalúa para hacer inteligible la realidad, para aprehender su significación, para comprender las maneras de resolver las situaciones planteadas, considerar sus errores como indicadores de problemas a descifrar, etc.” (Rizo, 2004, p. 24). Sin embargo, tal comprensión debe ser puesta al servicio de propósitos específicos para resultar efectiva, lo que conecta con las funciones o usos de la evaluación del aprendizaje.

Por influjo de las aportaciones iniciales de Scriven (1967), en la evaluación del aprendizaje se reconocen dos grandes usos o funciones: una social y otra pedagógica.

La **función social** de la evaluación se basa en comparar a las/os estudiantes entre sí o con respecto a una norma externa, para clasificarlas/os y tomar decisiones sobre su calificación y promoción académica (J. Vera & Esteve, 2001). Se trata de un uso o propósito corroborativo (Santos, 2010b) o apreciativo (Rizo, 2004) que determina logros o resultados de aprendizaje. Se corresponde con la evaluación sumativa, aquella presente desde la emergencia de la evaluación educativa, y conceptualizada formalmente en el paso de la segunda a la tercera generación evaluativa.

Como se detalla en el próximo capítulo, el uso sumativo es el que mejor representa la praxis evaluativa tradicional en los niveles universitarios y no universitarios. Sin embargo, los recelos que suscita su preeminencia son frecuentes y generales, en especial, en las dos últimas décadas. Así por ejemplo, propósitos típicamente sociales, como la medición, la clasificación o la discriminación han sido calificados como pobres (Santos, 2003, 2010b) o de alto riesgo (Knight, 2005), en referencia a la escasa contribución que realizan a la mejora y al desequilibrio entre su fiabilidad e impacto.

Harlen (2005, 2006) delimita dos modalidades de evaluación sumativa, dependiendo de la audiencia y la utilidad a la que sirva la función social desempeñada: la evaluación sumativa interna (dirigida a adoptar e informar las decisiones acreditativas en la propia institución formativa y ante el alumnado) y la externa (llamada a realizar una certificación

más allá de la propia institución). Adoptando los matices que aporta dicha clasificación, la tabla 13 recoge algunos de los propósitos específicos que más frecuentemente acompañan a la función social en la evaluación del aprendizaje.

Tabla 13

Propósitos específicos vinculados a la función social de la evaluación del aprendizaje

FUNCIÓN SOCIAL DE LA EVALUACIÓN: PROPÓSITOS ESPECÍFICOS	
INTERNO	EXTERNO
• Comprobar el grado de consecución de los objetivos para adoptar e informar decisiones relacionadas con la calificación (parcial o final), la certificación del aprendizaje o la promoción. • Proporcionar al alumnado un estímulo externo para motivar el trabajo y el aprendizaje. • Ejercer de instrumento para el control interno de responsabilidad e imputabilidad en los resultados formativos.	• Proporcionar datos para la transferencia, el reconocimiento y la certificación de los logros individuales de aprendizaje a efectos académicos y/o profesionales. • Proporcionar datos para la investigación educativa y/o procesos de evaluación externa. • Ejercer de instrumento para el control externo de responsabilidad e imputabilidad en los resultados formativos.
• Discriminar y clasificar resultados de aprendizaje y trayectorias académicas de calidad desigual, en función del grado de consecución en los objetivos de aprendizaje. • Convertir las evidencias de aprendizaje constatadas en una calificación que refleje el grado de consecución de los objetivos previstos. • Ejercer de filtro o mecanismo de selección en función de los resultados de aprendizaje evidenciados (mejores y peores trayectorias, satisfacción de requisitos mínimos, etc.). • Proporcionar datos para determinar el impacto y la eficacia del proceso de enseñanza y aprendizaje.	

Nota: La clasificación ha sido realizada ampliando las aportaciones iniciales de varias fuentes (Alegre & Villar, 2006; Casanova, 1998; Gibbs, 2003; Harlen, 2003, 2006; Patton, 1990; Rizo, 2004; M. P. Romero, 2004; Santos, 1993; Zuñiga, 1997). Los propósitos colocados en la fila inferior de la tabla son aplicables tanto a la evaluación sumativa interna, como a la externa.

Cualquiera de los propósitos específicos reseñados para la función evaluativa social, adoptan una orientación retrospectiva o hacia atrás como hacen notar Stufflebeam (1971) y Stake (2006). Sin embargo, dicha mirada no tiene por qué coincidir con el final del proceso formativo como ha venido sucediendo tradicionalmente (Brown, 2003; Patton, 1990; M. P. Romero, 2004). Ante la clásica identificación de la evaluación sumativa como final, cada vez son más las voces que señalan la proliferación de prácticas acreditativas en sistemas de evaluación continua (Bilbao, 2010; Margalef, 2009; Monereo & Castelló, 2009b; Sabirón & Arraiz, 2013). En tales circunstancias, el aprendizaje se desglosa en una colección variable de puntuaciones y fracciones de la calificación total que, a pesar de no llevarse a

cabo estrictamente al finalizar el proceso formativo, se efectúan desde una visión retrospectiva y con efecto social¹⁹.

Algo similar sucede a la hora de reparar en los instrumentos en los que se materializa la evaluación acreditativa; aunque tradicionalmente los exámenes hayan constituido el formato de evaluación sumativa por excelencia, cualquiera herramienta de evaluación puede servir a propósitos sociales internos o externos.

Pero además de su cometido social, la evaluación **desempeña una función pedagógica** centrada en la promoción y estimulación del aprendizaje discente. Esta segunda funcionalidad se caracteriza como rica (Santos, 2010b) y de bajo riesgo (Knight, 2005), orientada a una misión atributiva de comprensión de las causas, más que a la corroboración de resultados (Santos, 2003). Frente a la evaluación sumativa, centrada en la evaluación *del* aprendizaje, la funcionalidad pedagógica nos sitúa ante una evaluación *para* el aprendizaje (Arter, 2009; Harlen, 2003; Villardón, 2006; Wiliam, 2011). Su propósito principal no es la acreditación de los resultados, sino la adopción de una serie medidas o decisiones que favorezcan un mejor aprendizaje en un futuro inmediato y/o lejano. En consecuencia, adopta una orientación prospectiva o hacia delante (Stake, 2006; Stufflebeam, 1971), frente a la retrospección propia de la evaluación sumativa. La tabla 14 enumera alguna de sus manifestaciones más frecuentes:

Tabla 14
Propósitos específicos vinculados a la función pedagógica de la evaluación

FUNCIÓN PEDAGÓGICA DE LA EVALUACIÓN: PROPÓSITOS ESPECÍFICOS

- Proporcionar **información y orientación** sobre el proceso formativo, ya sea a nivel individual o grupal (retroalimentación, pautas de actuación para la mejora, etc.):
 - ✓ Implicar al alumnado en experiencias evaluativas en las que obtenga feedback o retroalimentación que le ayuden a comprender el estado actual de su aprendizaje.
 - ✓ Ayudar al alumnado a comprender e interiorizar las expectativas de logro hacia las que debe dirigir su aprendizaje y las condiciones para su evaluación.
 - ✓ Implicar al alumnado en experiencias evaluativas en las que obtenga sugerencias y pautas para mejorar su actuación en próximos intentos (proacción).
 - ✓ Promover situaciones de diálogo e intercambio comunicativo sobre el aprendizaje a desarrollar y su evaluación (entre alumnado y profesorado, entre estudiantes, etc.).
 - ✓ Estimular la reflexión sobre el aprendizaje a realizar y el desarrollo del proceso que media en su construcción.
- Proporcionar pautas para **regular o autorregular** el proceso de aprendizaje:
 - ✓ Generar en el alumnado una actividad adecuada y efectiva de aprendizaje.
 - ✓ Motivar el aprendizaje del alumnado.

¹⁹ Sin perjuicio de la lectura formativa que pueda realizarse en paralelo para la mejora de próximas actuaciones, cada resultado se determina con visión retrospectiva.

FUNCIÓN PEDAGÓGICA DE LA EVALUACIÓN: PROPÓSITOS ESPECÍFICOS

- ✓ Ofrecer al alumnado una guía efectiva para gestionar su tiempo de trabajo y esfuerzo a lo largo del proceso.
- ✓ Señalar los elementos centrales y nucleares del aprendizaje enfatizando su representación evaluativa.
- ✓ Orientar al alumnado en procedimientos o vías de trabajo efectivas por medio de las exigencias y condiciones evaluativas.
- Comprender el proceso de aprendizaje que lleva a cabo el alumnado (variables que interceden, momentos y factores críticos, aspectos nucleares, errores y dificultades habituales, necesidades asociadas a los diversos estilos de aprendizaje, obstáculos cognitivos y nociones erróneas que bloquean el avance, etc.) para **adaptar los procesos didácticos** en función de las necesidades reveladas:
 - ✓ Diagnosticar la situación de partida del alumnado con respecto al aprendizaje a efectuar.
 - ✓ Identificar los factores que mejor contribuyen al aprendizaje y las dificultades que obstaculizan su desarrollo.
 - ✓ Favorecer el seguimiento del proceso de aprendizaje en curso al alumnado y profesorado.

Nota: Elaboración propia a partir de las aportaciones iniciales de diversas fuentes (Alegre & Villar, 2006; Black y Wiliam, 2009; Brown, 2003b; Casanova, 1998; Gedye, 2010; Gibbs, 2003; Porto & Rodríguez, 2009; Rizo, 2004; M. P. Romero, 2004; Rotger, 1992; Sagasta & Bilbatua, 2009; Santos, 1993; Patton 1990).

Aunque en la literatura especializada se reconozcan de forma general los dos grandes usos señalados, la interpretación que se realiza de la función pedagógica no es unánime. Asumido de forma general el propósito de la mejora del aprendizaje, en su conceptualización se aprecian dos posturas diferentes con repercusión en la terminología empleada.

La postura más extendida en la literatura es aquella que identifica la función pedagógica de la evaluación con la evaluación formativa. Esta asimilación es deudora de la distinción entre evaluación sumativa y formativa planteada por Scriven (1967) y numerosas/os autoras/es recurren a ella (Banta, 2007; Black & Wiliam, 1998a, 1998b, 2009; Crossouard, 2010; Yorke, 2003; Winniger, 2005). En este caso, el término evaluación formativa se emplea en referencia a su función pedagógica y en contraposición a la sumativa.

La segunda postura diferencia dentro de la función pedagógica dos usos específicos, empleando de forma diferenciada los términos evaluación formadora y formativa (Parcerisa, 2010; Sanmartí, 2007, 2010; Sayós, 2010). De acuerdo con Bordas y Cabrera (2001),

la literatura sobre evaluación ha dejado bien clara la diferencia entre evaluación sumativa y formativa. Mientras que la evaluación sumativa orienta la toma de decisiones respecto a la calificación o la certificación, la evaluación formativa da luz

sobre ese indeterminado proceso de desarrollo. Pero es preciso avanzar, caminar hacia una evaluación formadora, es decir que arranque del mismo discente y que se fundamente en el autoaprendizaje. (pp. 31-32).

La evaluación formadora y la formativa comparten la motivación pedagógica de mejora. La diferencia de términos sólo denota el origen de dicha mejora: en la formativa es producto de una regulación externa (por lo general, a cargo del profesorado), mientras que en la formadora es interna (ejercida por el propio alumnado). De hecho, en ocasiones se habla directamente de evaluación reguladora y autorreguladora (Barberá, 1999; Coll & Onrubia, 1999; Jorba & Caselles, 1997) o de heterorregulación y autorregulación (O. Esteve, 2010).

- La *evaluación formativa* tiene por finalidad tomar decisiones que repercutan en la mejora del aprendizaje discente, ya sea adecuando el proceso didáctico (recursos, metodología, plazos y secuencias, etc.), u ofreciendo orientación o sugerencias de mejora directas al alumnado. Su principal destinatario es el profesorado, quien emplea la comprensión obtenida al servicio del aprendizaje discente (Stiggins, 2005a, 2005b; Stiggins & Chappuis, 2008).
- En la *evaluación formadora*, el profesorado “se debería centrar en compartir con los alumnos el proceso evaluativo, porque está comprobado que sólo el propio alumno puede corregir sus errores, dándose cuenta de por qué se equivoca y tomando decisiones de cambio adecuadas” (Sanmartí, 2007, p. 21). Se trata de una estrategia de evaluación dirigida a promover la autorreflexión y el control sobre el propio aprendizaje (Bordas & Cabrera, 2001; Cano & Ion, 2012; A. Pérez et al., 2009). Como facilitadora de la autorregulación, su presencia en la literatura ha cobrado mayor relevancia a raíz del proceso de convergencia donde, más que nunca, el alumnado está llamado a desarrollar su capacidad de aprendizaje a lo largo de la vida (I. Álvarez & Olivera-Smith, 2014; Hernández, Rosário, Cuesta, Martínez, & Ruiz, 2006, 2009; Villardón et al., 2013).

Esta diferencia de posturas dentro de la función pedagógica de la evaluación se traduce en un empleo desigual de los términos relacionados, alimentando la confusión en la literatura. Por lo general, la evaluación formativa aparece referida desde la visión amplia, como sinónima de evaluación pedagógica, bien diferenciada de la sumativa (social o acreditativa). Desde un criterio fundamentalmente práctico, esa es la postura general adoptada en el presente trabajo, aunque con ciertas excepciones para responder a sus

propósitos. Al tratar cuestiones precisas o cuando el detalle de la reflexión lo requiera, los conceptos de evaluación formadora y formativa se emplean de forma diferenciada, de acuerdo a sus propios matices. Por el contrario, cuando no aparezca referida frente a la formadora, la evaluación formativa se utiliza para designar la función pedagógica general.

A pesar de que las funciones social y pedagógica hayan sido presentadas por separado, ambas son actos de un mismo fin y, por tanto, no deben considerarse opuestas ni contrarias (Alegre & Villar, 2006; Brown, 2003; Edwards et al., 2008; Rizo, 2004; Stiggins, 2002). Como afirma Mateo (2000),

la evaluación es una acción conceptualmente única que, sin embargo, dependiendo del momento o de la necesidad, adopta funciones diversas. [...]. Ambas funciones se necesitan mutuamente, y responde a un planteamiento totalmente obsoleto el pensarlas como excluyentes; así desconocer el valor formativo de la evaluación sumativa nos parece una incongruencia e ignorar la aportación sumativa de la evaluación formativa un error. (p. 37).

La evaluación formativa no es suficiente (J. Chappuis et al., 2012). Las funciones social y pedagógica se complementan mutuamente, de modo que no debe plantearse una elección dicotómica entre las mismas (Banta, 2007; Knight, 2005; D. Salinas, 2002; Trillo, 2005). Stake (2006) suscribe esta postura, pero insiste en que la convivencia de ambos usos no lleve a su confusión: “La evaluación formativa y sumativa pueden tener lugar al mismo tiempo, pero conviene mantener separadas las funciones de mirar hacia delante con actitud formativa y mirar hacia atrás con actitud sumativa” (p. 63).

Desde esta última consideración, la evaluación sumativa, formadora y formativa no figuran entre las principales tipologías evaluativas expuestas en el próximo apartado, entendiendo que todas representan funciones de un mismo proceso y no categorías excluyentes.

5. Tipología evaluativa. Revisión y ampliación de las clasificaciones clásicas en la evaluación del aprendizaje

La ambigüedad terminológica que acompaña al fenómeno evaluativo se hace especialmente evidente al identificar y definir diferentes tipos de evaluación. A fin de poner algo de luz en esa confusión, el nuevo apartado presenta una propuesta revisada de clasificación.

El análisis de las tipologías que con mayor frecuencia se reseñan en la literatura, nos ha llevado a identificar 16 criterios que dan lugar a las principales tipologías evaluativas. Aunque no son las únicas, las modalidades recopiladas concentran las referencias habituales al tema y son las que mejor se ajustan a la evaluación del aprendizaje, aquella a la que hace referencia el presente trabajo. Expuestas de forma sintética en la tabla 15, su definición detallada se realiza en el **03 ANEXO**.

Tabla 15

Clasificación revisada de las principales tipologías aplicables a la evaluación del aprendizaje

EVALUACIÓN...	
MOMENTO	⇒ Inicial o diagnóstica ⇒ Procesual o estímulo ⇒ Final ⇒ Continua
CONTEXTO FÍSICO	⇒ Contexto académico ⇒ Contexto real (In situ) ⇒ Mixta
RECURSOS DE APOYO	⇒ Sin apoyo (a libro cerrado) ⇒ Apoyada ⇒ Mixta
AUTENTICIDAD	⇒ Descontextualizada ⇒ Mixta ⇒ Contextualizada • Ejecución/desempeño • Auténtica
NIVEL	⇒ Individual ⇒ Equipo ⇒ Grupal ⇒ Mixta

NORMOTIPO: CRITERIO DE EVALUACIÓN	⇒ Interna o idiográfica ⇒ Mixta ⇒ Externa o nomotética	• Normativa • Criterial
ATENCIÓN A LA DIVERSIDAD	⇒ Estereotipada ⇒ Diversificada	
ENFOQUE CURRICULAR	⇒ Disciplinar ⇒ Interdisciplinar	
FOCO DE ANÁLISIS	⇒ Analítica ⇒ Holística ⇒ Mixta	
TIPO DE EVIDENCIAS	⇒ Cuantitativa ⇒ Cualitativa ⇒ Mixta	
PROCEDIMIENTOS RECOGIDA	⇒ Monometódica ⇒ Plurimetódica	• Triangulada • No triangulada
DIVERGENCIA DE RESULTADOS	⇒ Convergente ⇒ Divergente ⇒ Mixta	
AGENTES EVALUACIÓN	⇒ Evaluación externa ⇒ Evaluación mixta ⇒ Evaluación interna	• Heteroevaluación • Autoevaluación • Coevaluación • Mixta
AGENTES DE CALIFICACIÓN	⇒ Calificación externa ⇒ Calificación mixta ⇒ Calificación interna	• Heterocalificación • Autocalificación • Co-calificación • Mixta
PARTICIPACIÓN DISCENTE	⇒ No participativa ⇒ Participativa	
DISTRIBUCIÓN DEL PODER	⇒ Autocrática ⇒ No autocrática	

Nota: Elaboración propia a partir de la revisión e integración de aportaciones parciales recopiladas de diversas fuentes (I. Álvarez, 2008a; AR@HKU, 2009b; Batallero, 2004; Biggs, 2005; Blanco, 2008a; Boud, Cohen, & Sampon, 1999; Bretones, 2008; A. Brew, 2003; Brown, Bull, & Pendlebury, 1997; Capllonch et al., 2009; Casanova, 1998; Castejón, Capllonch, González, & López, 2009; Durán, 2009; Engel & Bustos, 2009; Escamilla, 2010; Knight, 2005; V. M. López, 2011, 2012; Mateo, 2000; Meyer, 1992b; Monereo, 2009; Mueller, 2004; A. Pérez et al., 2009; Popham, 1980; Smither & Siso, 2007; Tejedor, 1999; Torrance, 1995; Trillo, 2005; Vu & Dall'Alba, 2007).

La tipología evaluativa resultante no es excluyente. Los criterios expuestos son compatibles y, en la mayor parte de los casos, cabe una modalidad mixta que combina las diferentes tipologías dentro la misma clasificación. La elección entre las diversas alternativas que ofrece este amplio abanico dependerá de las necesidades y circunstancias particulares que concurren en cada caso.

6. Conclusión: la evaluación del aprendizaje de calidad

El recorrido efectuado a lo largo de los anteriores apartados pone de manifiesto la complejidad natural que acompaña al proceso evaluativo. La coexistencia de generaciones, paradigmas y enfoques diferentes en su interpretación representa sólo una primera muestra de ello. Incluso unificada una determinada posición de partida, la complejidad de la evaluación se sigue haciendo presente en la cantidad y variedad de elecciones que exige su puesta en práctica: en qué momento evaluar, con qué propósito, con qué herramientas, en qué tipo de situación, a cargo de qué agentes, etc.

Para posicionarse entre todas estas opciones, las/os responsables del diseño evaluativo apoyan sus decisiones en dos tipos de referente: los de naturaleza práctica y los teóricos.

- Los factores prácticos de la evaluación representan las condiciones más directas e inmediatas que intervienen en su desarrollo. Las decisiones adoptadas deberían atender, al menos, a tres tipos de referentes presentados con mayor detalle en el capítulo quinto:
 - ✓ *Referentes pedagógicos o curriculares.* Las decisiones vinculadas a la evaluación deben responder al currículum de referencia, en especial a las competencias y objetivos previstos en el plan de estudio y a la estrategia metodológica perfilada para su desarrollo (Biggs, 2005; Biggs & Tang, 2011; Combs et al., 2008; L. Prieto, 2004).
 - ✓ *Referentes contextuales.* La evaluación no debe desprenderse de los rasgos propios del contexto de su puesta en práctica. En concreto, es importante que atienda a las particularidades del alumnado o al *perfil discente* de referencia (Knight, 2005; V. M. López, 2006, 2009b). Las decisiones

adoptadas también deben garantizar el cumplimiento de la *normativa evaluativa* dispuesta (Boud & asociados, 2010; Centre for the Study of Higher Education- CSHE, 2002; European Association for Quality Assurance in Higher Education- ENQA, 2011; Quality Assurance Agency for Higher Education- QAA, 2014) y plantear prácticas realizables en las condiciones de aplicación previstas (número de estudiantes, recursos disponibles, etc.) (McMillan, 2000).

- ✓ *Referentes metodológicos.* Entendida como indagación, la evaluación debe preservar una serie de medidas y cautelas que garanticen la adecuación (validez) y precisión (fiabilidad) de las conclusiones que alcance. Como en todo proceso de investigación, tampoco puede dejar pasar por alto los derechos que asisten a personas involucradas, esto es, las *implicaciones éticas* de las decisiones adoptadas.
- La teoría de la evaluación representa el *debería ser* o la visión sostenida sobre la evaluación de calidad. Se concreta en el conjunto de interpretaciones, concepciones y creencias sobre los rasgos deseables de la evaluación y adopta formas diversas, incluso dentro de un mismo paradigma evaluativo. A menudo permanece en el inconsciente, pero esa visión particular condiciona poderosamente las decisiones y actuaciones vinculadas a las distintas fases de la evaluación.

Centrada en la competencia evaluativa docente, en la presente investigación resulta especialmente relevante definir la teoría de la evaluación de partida. No existe una única visión al respecto, de modo que, en torno a la evaluación convergen preferencias e interpretaciones dispares. Por ese motivo, antes de delimitar los indicadores que ha de evidenciar en la práctica evaluativa el profesorado, se hace necesario declarar aquellos rasgos y condiciones que se asocian a la evaluación del aprendizaje de calidad.

A lo largo de este segundo capítulo ya se han efectuado algunas delimitaciones iniciales sobre la teoría de la evaluación desde la que se enfoca nuestra aportación. La explicitación de una definición de partida y su adscripción a un paradigma evaluativo determinado proporciona una primera visión general. Partiendo de la misma, el capítulo cuarto presenta en detalle y justifica el modelo de práctica evaluativa deseable al que se orienta este trabajo, estableciendo subprocesos, criterios e indicadores para su definición. El resultado es el modelo ITC (Integrado Teórico Complejo) para la evaluación del aprendizaje en los grados vinculados a la formación básica de maestras/os. La propuesta de sistematización de la competencia evaluativa docente expuesta en el capítulo quinto sólo puede representar una referencia útil en caso de compartir tales premisas.

7. El capítulo a grandes rasgos

Este segundo capítulo ha servido al objetivo de establecer un marco de referencia conceptual para profundizar en la evaluación del aprendizaje, ante la divergencia de interpretaciones y la polisemia imperante.

La presente investigación se aproxima al *paradigma de la indagación y adopta la conceptualización de la evaluación para el aprendizaje* como principal referencia dentro del mismo. Situada en el enfoque de la evaluación alternativa, dicha conceptualización ofrece un marco de referencia amplio en el que poder integrar otras conceptualizaciones asociadas a la perspectiva indagadora (evaluación participativa, sostenible, basada en competencias, auténtica y de la ejecución o desempeño).

Partiendo de tal posición, definimos *la evaluación como el proceso sistemático de recogida y análisis de información y evidencias sobre una serie de criterios delimitados de antemano, para comprender un objeto, fenómeno o situación determinada y actuar en consecuencia*. Esta delimitación recoge los aspectos clave identificados en definiciones de la literatura clásica y reciente y en otros trabajos que han tratado de sintetizarlos:

- El proceso de evaluación debe procurar el recorrido consciente y sistemático de *cuatro grandes fases o etapas*: la planificación, la puesta en práctica, el establecimiento de conclusiones y la metaevaluación. Su desarrollo incluye más de veinte decisiones o actuaciones interrelacionadas, pero no se trata de un esquema de actuación rígido.
- Aunque puede atender a objetos diversos, en la presente investigación la evaluación se orienta a la comprensión del aprendizaje de las/os estudiantes. Este aprendizaje se asimila con el rendimiento académico, entendido como *el desarrollo progresivo que realiza el alumnado de las competencias y objetivos y previstos, según se manifiesta en las actividades evaluativas acreditativas y pedagógicas desarrolladas*. Esta visión evita la equiparación entre rendimiento y calificación, contemplando el valor de la evaluación como medio didáctico para fortalecer el aprendizaje, y no sólo para acreditarlo.
- Satisfecha la motivación inicial de la comprensión, la evaluación del aprendizaje debe desencadenar en una actuación que puede ponerse al servicio de *finés sociales (evaluación sumativa o acreditativa) o pedagógicos (evaluación formativa y formadora)*. El uso de las conclusiones de la evaluación en un sentido u otro, dependerá de lo que haya sido establecido en la fase de diseño. Ambas son

funciones de la evaluación y desempeñan misiones complementarias en la estimulación y acreditación del aprendizaje.

Concebida desde tales premisas, la evaluación del aprendizaje puede dar lugar a manifestaciones muy diversas. La revisión y reordenación de las tipologías clásicas, nos ha llevado a identificar 16 criterios a partir de los cuáles clasificar las principales modalidades de evaluación empleadas en el contexto formativo de referencia. La definición de cada una de ellas, así como de los conceptos evaluativos clave que serán empleados en los apartados sucesivos, han sido trasladados al apartado de los anexos (**ANEXO 02 y 03**) para favorecer el seguimiento unívoco de la memoria de investigación.

CAPÍTULO 3.

LA EVALUACIÓN DEL APRENDIZAJE DE LAS/OS ESTUDIANTES UNIVERSITARIAS/OS EN EL SISTEMA UNIVERSITARIO ESPAÑOL

Calificadas como paradojas o contradicciones (Litwin, 1998; Santos, 1988, 1993, 1999), desequilibrios (Villardón, 2006), patologías evaluativas (Batalloso, 2004) o esquizofrenias (V. M. López, 2004), las alusiones a las deficiencias y limitaciones de los sistemas tradicionales de evaluación del aprendizaje en la universidad son frecuentes y abundantes. Los múltiples diagnósticos llevados a cabo coinciden en describir un sistema de evaluación del aprendizaje con marcadas deficiencias, que no parece complacer a ninguna de las partes directamente involucradas.

Más que un *reto* general de mejora, en el presente trabajo la evaluación del aprendizaje discente se eleva a la categoría de *problema*, por el cúmulo de limitaciones y dificultades que tradicionalmente le acompañan. Siguiendo a Zabalza (2006) o Crossouard (2010), podríamos concluir que no nos enfrentamos al clásico problema de la evaluación en la universidad, sino al gran problema del aprendizaje universitario.

Afrontar con garantías de éxito la mejora de la evaluación del aprendizaje en la universidad exige, en primer término, identificar y analizar cuáles son aquellos rasgos y prácticas que caracterizan su planificación y desarrollo. Dicha descripción constituye el punto de partida del nuevo capítulo, en el que se añaden consideraciones complementarias en torno al fenómeno evaluativo hasta ofrecer un resumen completo del estado de la cuestión. Integrando las aportaciones realizadas desde la literatura y la investigación específica, en las siguientes páginas se realiza el recorrido representado en la figura 11:

- Los rasgos que caracterizan a las prácticas evaluativas tradicionales en el sistema universitario español.
- Los principales efectos que éstas ocasionan, tanto en el colectivo discente, como en el profesorado.
- Las causas estructurales y profundas que se asocian a tales deficiencias.

- Las líneas principales de intervención y sugerencias que se formulan en aras a la mejora.
- En coherencia con los propósitos de la presente investigación, también se reserva un espacio para reflexionar sobre las prácticas evaluativas específicas que tienen lugar en las titulaciones orientadas a la formación básica de maestras/os.
- El capítulo finaliza con una síntesis de las claves fundamentales tratadas a lo largo de su desarrollo.

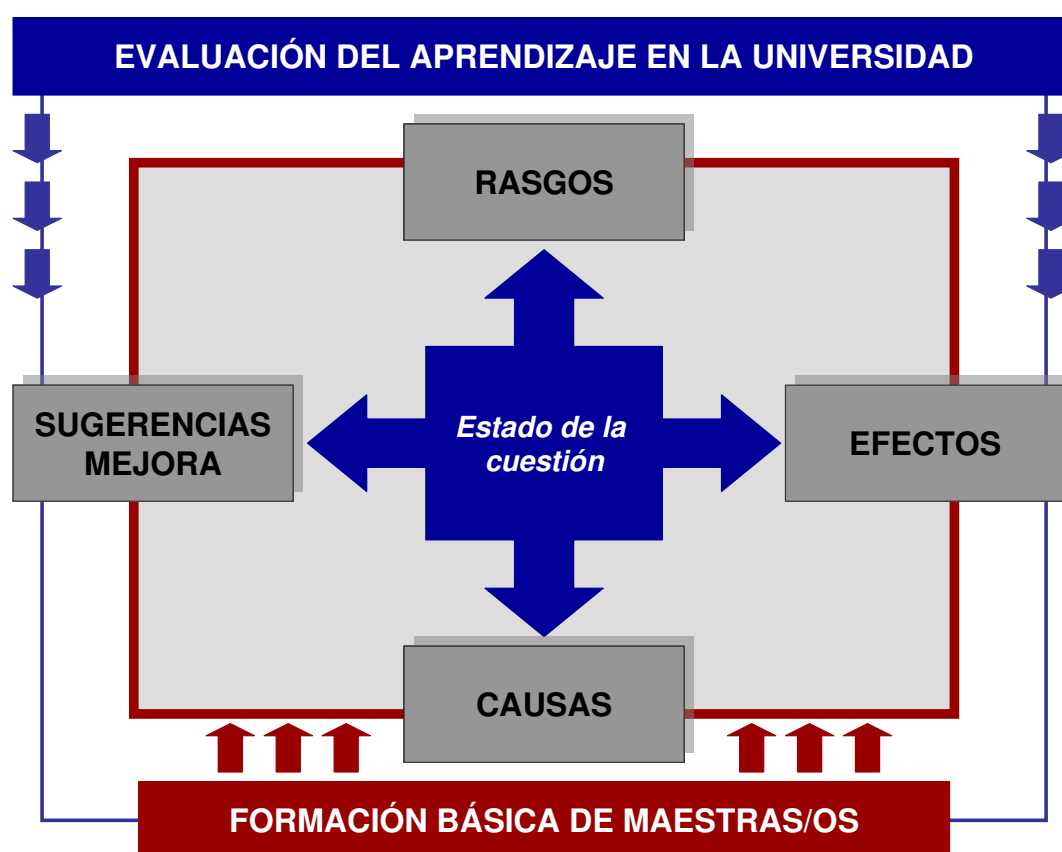


Figura 11. Evaluación del aprendizaje universitario: análisis del estado de la cuestión.

1. Deficiencias generales de los sistemas tradicionales de evaluación del aprendizaje universitario en España

En un marco universitario habituado a desenvolverse en la incertidumbre natural de los procesos de renovación y cambio, el problema de la evaluación del aprendizaje representa un viejo conocido, una asignatura pendiente nunca del todo resuelta (Bilbao,

2010). Ramsden (2003) afirma que, salvo en la mitología académica, nunca existió una época dorada para la docencia universitaria y la cuestión se vuelve mucho más oscura al tratar sobre la evaluación del aprendizaje, una de las competencias más complejas ligadas a ella. El profesorado puede llegar a dedicar a las tareas evaluativas entre un tercio y la mitad de su jornada (Stiggins, 1993/2012; Stiggins & Conklin, 1988). Sin embargo, la evaluación del aprendizaje representa una actividad pedagógica temida y estresante para estudiantes (Serrano, 2010; Stiggins, 1999) y la tarea más tediosa, molesta y desmotivadora para el profesorado (Atkinson & Lim, 2013; Trillo, 2005; Flórez, 1999).

La evaluación del aprendizaje se ha constituido como centro de muchos estudios e investigaciones que han pretendido llevar a cabo un diagnóstico de su presencia y efecto en los procesos formativos. Desarrollados en contextos y niveles educativos diversos, dichos estudios se desarrollan bajo el propósito común de analizar la praxis evaluativa, ya sea de forma global, o atendiendo específicamente a alguno de sus componentes.

La tabla 16 presenta algunos de los trabajos que han sido considerados para sintetizar la situación de partida en el tema que nos ocupa. Su clasificación tiene lugar de acuerdo a tres grandes criterios:

- El **ámbito geográfico** en el que ha sido desarrollado el análisis, diferenciando entre estudios de ámbito nacional (NAC) e internacional (INT). Por el contexto y los propósitos de la investigación en curso, se ha primado la consideración de los primeros frente a los segundos, basados en una realidad más lejana y menos representativa.
- La **titulación o ámbito formativo específico** en el que se ha centrado el análisis, distinguiendo entre los estudios desarrollados en el contexto de la formación básica de maestras/os o profesorado (FB) y en otras titulaciones (científicas, técnicas o humanas y sociales) (NO). Para sintetizar y describir las limitaciones y deficiencias evaluativas de las que da cuenta el presente capítulo se ha prestado atención preferente a las primeras.
- La **perspectiva** desde la que se ha realizado el análisis lleva a discriminar entre estudios basados en el punto de vista del profesorado (P), de las/os estudiantes (E), del alumnado egresado (AE), de superiores directos (SUP) y aquellos otros basados en el análisis de documentos (DOC) (revisión de normativas y marcos reguladores, programas, etc.). Como se aprecia en la propia tabla, estas fuentes de información no son excluyentes.

Aunque han aportado datos relevantes sobre la evaluación del aprendizaje, no se incluyen en este resumen estudios más amplios sobre la valoración global de la calidad o eficacia docente (V. Álvarez et al., 2011; V. Álvarez, García, & Gil, 1999; Arribas et al., 2010; Gutiérrez, Pérez, Pérez, & Palacios, 2011; L. Prieto, 2007; Rué, 2009b). Tampoco trabajos de revisión o meta-análisis sobre aspectos concretos de la evaluación (Bretones, 2008; Dochy, Segers, & Sluymans, 1999; Evans, 2013; Ferguson, 2011; Jonsson, 2013; McGarr & Clifford, 2013; Struyven, Dochy, & Janssens, 2005).

Tabla 16
Antecedentes en el análisis de la praxis evaluativa

	ÁMBITO		TITULACIÓN		PERSPECTIVA/S				
	NAC	INT	FB	NO	P	E	DOC	SUP	AE
Baartman, Gulikers, & Dijkstra, 2013		X		X	X	X			
C. Brew et al. (2009)		X	X		X	X			
Cano & Ion (2012)	X			X	X				
Clipa & Duta (2010)		X		X		X			
M. Fernández (1989)	X			X		X			
Flórez (1999)	X			X		X			X
Gil-Flores (2012)	X			X		X			
Gutiérrez et al. (2013)	X		X		X	X			X
Hamodi & López (2012)	X		X			X			X
Hortigüela et al. (2015)	X		X			X			
Ibarra & Rodríguez (2010)	X			X			X		
V. M. López & Palacios (2012)	X		X			X			
MacLellan (2001)		X		X	X	X			
L. F. Martínez et al. (2012) ^a	X		X		X	X			
N. Martínez et al. (2012)		X		X		X			
Martínez-Mínguez & Flores (2014)	X		X		X				X
McDowell (1995) ^b		X		X		X			
Nieto (2000)	X			X	X				
Olmos & Rodríguez (2010)	X			X	X				
Padilla et al. (2010)	X			X			X		
Palacios & López (2013)	X		X		X				
Palacios et al. (2013)	X		X		X		X		
M. L. Pérez et al. (2000)	X		X		X		X		
Porto (2006) ^c	X			X		X		X	
Porto & Rodríguez (2009) ^d	X			X		X			

	ÁMBITO		TITULACIÓN		PERSPECTIVA/S				
	NAC	INT	FB	NO	P	E	DOC	SUP	AE
Ricoy & Fernández-Rodríguez (2013) ^e	X			X		X			
G. Rodríguez et al. (2012)	X			X	X	X	X		
G. Rodríguez (1998)	X			X	X				
Ruiz-Gallardo, Ruiz, & Ureña (2013)	X		X		X	X			
Sabaté, Marzo, Miró, Martínez, & Añaños (2008)	X			X	X				
Smitter (2006)		X	X			X			
Smitter & Siso (2007)		X	X				X		
Stiggins & Conklin (1988)		X	X				X		
Stiggins (1998)		X	X				X		
Tejedor (1999)	X			X	X	X			
Trillo & Porto (1999) ^f	X			X		X			
Vargas (2008)		X		X	X	X			

Nota: En la columna que detalla el ámbito (NAC) es nacional e (INT) internacional; en la columna relativa a la titulación se diferencia entre estudios desarrollados en la formación básica de maestras/os o profesorado (FB) o en otras titulaciones; en la columna perspectiva se detalla la fuente considerada en el estudio: (P) Profesorado; (E) Estudiante; (DOC) Documentos; (SUP) Superiores directos; (AE) Alumnado egresado.

^a El estudio se centra específicamente en la formación inicial de profesorado de Educación Física.

^b Sobre dichos estudios se realiza una reelaboración posterior en McDowell y Sambell (2003).

^c Sobre dicho estudio se realiza una reelaboración posterior en Porto (2008).

^d Plantea modificaciones sobre el cuestionario empleado en Trillo y Porto (1999).

^e El estudio se lleva a cabo en el contexto del Máster Secundaria.

^f Sobre este mismo estudio ofrecen elaboraciones posteriores Trillo y Porto (2002) y Porto (2005a, 2005b).

A pesar de la diversidad de enfoques y planteamientos, los análisis y revisiones realizados sobre la evaluación del aprendizaje revelan conclusiones con un marcado nivel de uniformidad y coincidencia. Desde esta apreciación, interpretamos:

- Las grandes deficiencias de la evaluación del aprendizaje se repiten en los distintos niveles educativos, aunque la gravedad de las manifestaciones es más acusada en unos que en otros. De acuerdo con D. Salinas (2002) y Santos (2003), en la medida en que aumenta el nivel educativo, aumentan también las prácticas disfuncionales vinculadas a la praxis evaluativa, en una tendencia sujeta a distintas interpretaciones posibles (diferentes condiciones organizativas y estructurales, particularidades del perfil discente, bagaje formativo distinto en el profesorado de unos niveles y otros, etc.).

- Las audiencias más directamente involucradas en la evaluación del aprendizaje (estudiantes y profesorado) manifiestan una visión similar del estado de la misma. Con los matices y diferencias connaturales al rol que cada uno desempeña en el proceso (V. Álvarez, García, Gil, Romero, & Correa, 2002), las limitaciones y deficiencias señaladas muestran un nivel de concordancia razonable.
- Los estudios centrados en el diagnóstico de los sistemas de evaluación del aprendizaje universitario, revelan síntomas muy similares a nivel nacional e internacional, aunque las condiciones y planteamiento de partida difieran de unos países a otros.
- Los trabajos que dan cuenta de buenas prácticas evaluativas, hacen referencia a la valoración de experiencias puntuales de innovación (Haigh, 2007; Ureña et al., 2009; Vu & Dall'Alba, 2007; Winniger, 2005). No hemos dado con ningún estudio que revele prácticas evaluativas satisfactorias y saludables de forma generalizada, con independencia del contexto geográfico y formativo y de la fuente consultada.
- Estudios desarrollados en un intervalo cronológico bastante dilatado, revelan deficiencias similares en lo que respecta a la evaluación del aprendizaje universitario. Situados en el contexto español, uno de los primeros diagnósticos de impacto sobre el tema fue el de M. Fernández (1989), que dejaba constancia del predominio de prácticas evaluativas finalistas, basadas en la realización de exámenes escritos y con intervención exclusiva del profesorado. Transcurridas dos décadas, trabajos recientes no dibujan una situación muy diferente; aunque la implantación del EEES haya favorecido ciertos cambios y avances, en esencia, la evaluación sigue presentándose como un elemento prioritariamente final, frecuentemente asociado a un examen escrito, que determina la calificación y evoca sensaciones negativas (Ricoy & Fernández-Rodríguez, 2013).

1.1 Descripción de síntomas.

La evaluación del aprendizaje del alumnado universitario español, se ve afectada por una amplia colección de síntomas de disfuncionalidad ampliamente reportados en las fuentes nacionales consultadas. Todos estos rasgos guardan entre sí una estrecha relación de interdependencia por la que se condicionan mutuamente (ver la tabla 17):

Tabla 17

Síntomas de disfuncionalidad en la evaluación del aprendizaje universitario

La evaluación en la universidad...

- a) Se centra casi de manera exclusiva en la figura discente.
- b) Es unidireccional: heteroevaluación docente sobre el aprendizaje del alumnado.
- c) Se centra en la consideración de productos (fundamentalmente individuales).
- d) Es básicamente final.
- e) Presenta un carácter predominantemente acreditativo sumativo.
- f) Se realiza primordialmente a partir de exámenes.
- g) Se lleva a cabo por medio de instrumentos inadecuados de evaluación.
- h) Incurrir en una excesiva simplificación del aprendizaje cuya valoración se pretende.
- i) Se muestra independiente de la enseñanza y el aprendizaje.
- j) Se muestra desalineada de los objetivos y la metodología.
- k) No responde con contundencia a las exigencias de explicitación y transparencia.
- l) Ofrece un feedback deficiente, insuficiente y poco frecuente.
- m) Es homogénea o uniforme.
- n) Favorece la adopción de enfoques estratégicos y superficiales de aprendizaje.
- o) Es fundamentalmente convergente.
- p) Es atomizada, parcelada y disciplinar.
- q) Se produce de forma descontextualizada.
- r) Tiene una dudosa capacidad de predicción sobre el desempeño profesional real.
- s) Responde más a condiciones prácticas, normativas y administrativas que a pedagógicas.
- t) No se somete a revisión o metaevaluación.
- u) Ofrece señales evidentes de descoordinación profesional al nivel de titulación.
- v) No complace a docentes, a discentes, ni a audiencias externas.

Finalizada su exposición preliminar, el próximo apartado sugiere una síntesis que agrupa esta enumeración de deficiencias en torno a cinco núcleos (ver tabla 18).

a) *La evaluación del aprendizaje en la universidad se centra casi de manera exclusiva en la figura discente.*

Según revelan los resultados de los principales estudios y diagnósticos consultados, la evaluación del aprendizaje en la universidad se centra casi de forma exclusiva en la figura y el papel discente (Monescillo & Toscano, 2007; Porto & Rodríguez, 2009; Smitter, 2006). Como señalan V. M. López (2004) y Santos (1999, 2009), esta artificial reducción puede conducir a procesos atributivos gratuitos, que explican

todas las deficiencias y carencias identificadas en el aprendizaje desde un único foco de atención: el alumnado.

b) La evaluación del aprendizaje en la universidad es unidireccional de modo que se practica la heteroevaluación docente sobre el aprendizaje del alumnado.

A pesar de que en los últimos años se acusa un cierto avance (I. Álvarez, 2008a; Glasner, 2003), las experiencias de autoevaluación y coevaluación en el contexto universitario siguen siendo puntuales y secundarias. Así se evidencia considerado el testimonio de estudiantes y docentes (G. Rodríguez & Ibarra, 2012; G. Rodríguez et al., 2012) y al analizar los programas y marcos reguladores para la praxis evaluativa (Padilla et al., 2010).

Por lo general, el alumnado no participa de la evaluación (a través de estrategias alternativas como la autoevaluación, la evaluación por compañeros o la co-evaluación), sino que, por un lado, es receptor de las consecuencias de la evaluación -la calificación, la superación de la asignatura y la obtención del título- y por otro, disfruta de una serie de derechos que se vinculan al régimen académico derivado de la matrícula en las asignaturas. (Ibarra & Rodríguez, 2010, p. 394).

La pasividad propiciada desde esta concepción evaluativa no es consistente con los planteamientos del EEES (Bretones, 2008; Sayós, 2010), ni con las finalidades pedagógicas de autorregulación y transformación ética de la sociedad propias de las instituciones universitarias (Rué, 2007). Como señalan Trillo y Porto (2002), esta incongruencia entre el papel central del alumnado en la formación y su escasa implicación evaluativa, excede el ámbito de responsabilidad de las universidades, ya que las agencias de evaluación y/o acreditación de la calidad institucional tampoco contemplan la participación discente en la evaluación entre sus orientaciones y directrices²⁰.

²⁰ Así se constata en el capítulo cuarto, al analizar las directrices evaluativas de aplicación estatal formuladas desde la ENQA (2009) o el programa AUDIT (ANECA, AQU, & ACSUC, 2007a, 2007b), así como otras originadas en el contexto internacional (Quality Assurance Agency- aprendizaje, 2014).

c) La evaluación del aprendizaje de las/os estudiantes universitarias/os se centra en la consideración de productos (fundamentalmente individuales), llevando a desconsideración los procesos que median en su consecución.

La evaluación del trabajo realizado en grupos o equipos está sujeta a importantes beneficios potenciales. El enriquecimiento de la calidad del aprendizaje, el desarrollo del sentido de la responsabilidad, la estimulación de competencias profesionales genéricas o la disminución de la sobrecarga docente que genera la evaluación son algunos de los más reseñados (Chan, 2010). Sin embargo, la evaluación del aprendizaje atiende prioritariamente a los productos, desconsiderando el desarrollo de los procesos que median en su consecución (Porto & Rodríguez, 2009; Smitter, 2006; Smitter & Siso, 2007; Villardón, 2006). La mayor parte de las veces, estos productos son de autoría individual (Race, 2003; Santos, 1999). De acuerdo con D. Salinas (2002) esta tendencia da lugar a una cultura evaluativa individualista desde los niveles inferiores del sistema educativo.

d) La evaluación del aprendizaje de las/os estudiantes en la universidad es básicamente final.

Unido a la consideración casi exclusiva de los productos, la evaluación del aprendizaje universitario es básicamente final. Los diagnósticos y revisiones realizados sobre el tema describen prácticas evaluativas puntuales y finalistas, particularmente aquellos publicados antes de que se iniciase el proceso de convergencia (Flórez, 1999; Glasner, 2003; I. González, 2003; Nieto, 2000; G. Rodríguez, 1998; Tejedor., 1999; Trillo & Porto, 1999). Estudios más recientes sugieren un cierto avance en la proliferación de formatos de evaluación continua (Coll, Rochera, Mayordomo, & Naranjo, 2007; Ibarra & Rodríguez, 2010; Olmos & Rodríguez, 2010; Padilla et al., 2010; Villa et al., 2013), presumiblemente por la consolidación progresiva del planteamiento formativo sujeto al EEES. Aun así, la plena integración de la evaluación en el proceso formativo, sigue presentándose como área de mejora recurrente en la enseñanza universitaria (V. M. López, 2006, 2009a; V. M. López & Palacios, 2012).

Las evaluaciones llevadas a cabo en un único momento no captan el dinamismo de la construcción del aprendizaje desde la óptica constructivista (Shepard, 2000). Sin embargo, los intentos por desarrollar dispositivos evaluativos que atiendan a los avances y dificultades del proceso, a menudo se convierten en una sucesión de pruebas que conducen a la nota final (Biscarri, Filella, & Jovés, 2006; Murphy, 2006; Sabirón & Arraiz, 2013; Stobart, 2010). Monereo (2003) propone la analogía entre la

evaluación continua así entendida y una especie de pista americana. Estos sistemas basan la calificación del aprendizaje en la suma o el promedio de varias partes que no necesariamente reflejan el grado de consecución de los objetivos (Arribas et al., 2010), además de generar descuelgues, alimentar tendencias obsesivas (Rizo, 2004) y debilitar la autoestima de las/os estudiantes más débiles (Monereo & Castelló, 2009b). En este contexto, la evaluación inicial o diagnóstica sigue siendo infrecuente y prácticamente anecdótica en la universidad (Clipa & Duta, 2010; A. Gargallo, 2009).

e) La evaluación del aprendizaje de las/os estudiantes en la universidad presenta un carácter predominantemente acreditativos o sumativo.

Además de final, la evaluación del aprendizaje en la universidad responde básicamente a la finalidad acreditativa (Banta, 2007; MacLellan, 2001). Aunque se admite que las funciones sumativa, formativa y formadora de la evaluación no son excluyentes ni contrapuestas (J. Chappuis et al., 2012; S. Chappuis & Chappuis, 2007; Knight, 2005; Winniger, 2005), el predominio de la primera es la conclusión que más se reitera en los estudios sintetizados en la tabla 16. Se evalúa para calificar, no para orientar el aprendizaje (Àgencia per a la Qualitat del Sistema Universitari a Catalunya - AQU, 2003; Tejedor, 1999) y la atención se centra en la medición, no en la estimulación (Birenbaum et al., 2006; Carless, 2002, 2007; Stiggins, 2002; Stiggins & Chappuis, 2005). Siguiendo a Andreu-Andrés y Labrador-Piquer (2011) esta premisa representa el denominador común en las diversas conceptualizaciones que sostiene el profesorado en torno al fenómeno evaluativo.

Este predominio de la finalidad acreditativa se traduce en una desconsideración del carácter preventivo de la evaluación y en una artificial limitación de su potencialidad didáctica.

En ningún momento se hace mención a la evaluación como un mecanismo para la mejora del aprendizaje. Más bien al contrario, el análisis de la normativa nos lleva a pensar que no siempre se diferencia entre evaluación y calificación, y que esta última es esencial para discernir si un alumno ha superado o no un período formativo, cuestión importante a cuya resolución se orienta todo el sistema de evaluación. (Padilla et al., 2010, p. 11).

La preponderancia de la función acreditativa en la evaluación deja tras de sí consecuencias importantes, entre las que consideramos prioritarias las tres siguientes:

- La sobrerrepresentación de la función sumativa de la evaluación lleva a subrayar la calificación del aprendizaje académico por encima de su valor de uso (Jorba & Sanmartí, 2000; Race, 2003; Stobart, 2010). En consecuencia, el deseo y el disfrute del aprendizaje se desvanece bajo la presión por el resultado (Santos, 2001).
- El énfasis en la función sumativa penaliza y sanciona el error: lo convierte en algo negativo y lo desestima como fuente de información (J. Chappuis, 2014). La evaluación se centra en la detección de los errores (Brown, 2003b; Huxham, 2007; Wootton, 2002) y la preeminencia de la función acreditativa no deja espacio para la experimentación y la equivocación (Bilbao, 2010). En consecuencia, a pesar de las grandes potencialidades didácticas y pedagógicas atribuidas al error²¹ se generan conductas de evitación entre los agentes evaluados y de penalización por parte de los evaluadores. Batalloso (2004) y Monescillo & Toscano (2007) hablan de una función sancionadora, punitiva y selectiva de la evaluación en este sentido.
- La evaluación sumativa y final característica de la formación universitaria, envía un mensaje contrario a la política de formación a lo largo de la vida.

Anima a los estudiantes y los empleadores a centrarse en un resultado final percibido como punto de cierre, en vez de abrirlos a una amplia y diversa gama de modalidades y niveles de logro, como parte de un proceso continuo de aprendizaje que continuará más allá de la consecución de su título. (Burgess, 2007, p. 7).

Pero la más grave de las repercusiones que propicia esta situación es que, de hecho, la mayor parte del alumnado no considera que los sistemas evaluativos experimentados representen una ayuda para su aprendizaje (Hamodi & López, 2012; Tejedor, 1999). Frente a ello, Coronel (2007) concluye:

²¹ Descritas, entre otras/os muchos por Sanmartí (2007), Barth (2004), Bain (2007), Torre (2008) o Villa (2004).

La evaluación no es un sistema para detectar lo que lo que el estudiante no sabe, sino para ayudarle a sentar lo que sabe y a incentivarle para que aprenda lo que desconoce. La evaluación no es tampoco una herramienta de autoridad en la mano del profesor. La evaluación es una ocasión para profundizar en los contenidos de la materia. (p. 30).

f) *La evaluación del aprendizaje de las/os estudiantes en la universidad se realiza primordialmente a partir de exámenes.*

Birenbaum (1996) habla de la cultura evaluativa de la prueba para referirse a otro rasgo evaluativo ampliamente extendido en la literatura: la hegemonía del examen como técnica de recogida de evidencias por excelencia en el contexto universitario.

Al igual que en la continuidad de la evaluación, estudios publicados en los últimos veinte años sugieren una evolución positiva en este sentido. Frente a los primeros diagnósticos, en los que el examen aparecía casi como único procedimiento evaluativo (M. Fernández, 1989; Flórez, 1999; Nieto, 2000), investigaciones más recientes sugieren que el empleo de otras técnicas e instrumentos se extiende progresivamente (Cano & Ion, 2012; Olmos & Rodríguez, 2010; Padilla et al., 2010). Aun así, el protagonismo del examen sigue siendo una conclusión válida en nuestros días, por su mayor peso y presencia en los sistemas evaluativos empleados (Arribas et al., 2010; Gil-Flores, 2012; Gutiérrez et al., 2011; L. F. Martínez et al., 2012; Martínez-Mínguez & Flores, 2014; Ricoy & Fernández-Rodríguez, 2013).

Las investigaciones desarrolladas sobre el tema alertan de ciertos riesgos asociados a este predominio:

- La tradición evaluativa universitaria se centra fundamentalmente en la evocación y reproducción de aprendizajes conceptuales o declarativos y así se manifiesta en los exámenes empleados (V. Álvarez et al., 2002; Baartman et al., 2013; Gutiérrez et al., 2013; Ruiz-Gallardo et al., 2013). Por las dificultades que entraña su valoración, los componentes actitudinales del aprendizaje y las destrezas vinculadas a la práctica tienden a ser omitidos de las pruebas acreditativas (Brown, 2003a; S. Rodríguez, 2005; Villardón, 2006).
- Las demandas teóricas, reproductivas y convergentes predisponen al alumnado a la memorización y repetición del contenido, actuaciones poco compatibles con resultados de aprendizaje profundo (Castejón et al., 2009;

Dochy et al., 2002; Scouller, 1996, 1998). Sin embargo, por efecto del hábito, las/os estudiantes recelan y se muestran resistentes ante la aplicación de sistemas evaluativos más innovadores, en los que advierten amenazas al principio de justicia (Struyven et al., 2005) e inseguridad (Bevitt, 2015; Forés, Sánchez, & Sancho, 2014).

- El propio profesorado manifiesta dudas sobre la capacidad de los exámenes empleados para captar de forma válida y fiable los aprendizajes pretendidos (Arribas, 2012); a pesar de ello, su elección se justifica por la necesidad de atender a una cantidad importante de estudiantes y de garantizar la uniformidad de criterios y condiciones (M. L. Pérez et al., 2000; Tejedor, 1999).
- Las pruebas o exámenes pueden tener un efecto negativo en la motivación de las/os estudiantes (Harlen & Deakin, 2003; Stiggins, 1999, 2007). Sanmartí (2007) afirma que los exámenes sólo sirven de estímulo y motivación a quienes tienen éxito y Race (2003) que restringen el interés por aprender.

Pero estos riesgos o inconvenientes sobre los que se podría seguir profundizando, no son connaturales a los exámenes, sino que responden a una deficiencia más amplia y global expuesta a continuación.

g) La evaluación del aprendizaje de las/os estudiantes en la universidad se lleva a cabo por medio de instrumentos inadecuados de evaluación.

La literatura describe síntomas de inadecuación amplios y variados en la instrumentación puesta al servicio de la evaluación. Además de la excesiva simplificación del aprendizaje sugerida en el próximo rasgo, resultan especialmente destacables los presentados a continuación:

- Toda técnica o instrumento de evaluación cuenta con ventajas y limitaciones que conviene contrarrestar mediante procedimientos de triangulación (Leal, 2004). Sin embargo, la investigación específica da cuenta de evaluaciones monometódicas o monoinstrumentales (Batalloso, 2004; Glasner, 2003).
- Unido a lo anterior, en la universidad se emplea un repertorio de técnicas poco variado y muy restringido, que no permite demostrar al alumnado sus capacidades y logros con respecto a los objetivos de aprendizaje (Brown, 2003b; Erwin, 2003; McDowell & Sambell, 2003).

- La falta de reflexión y justificación en la selección de los instrumentos que se emplean en la evaluación del aprendizaje representa otra crítica frecuente (AQU, 2003; Joint Committee on Standards for Educational Evaluation-JCSEE, 2007; Padilla et al., 2010). Las revisiones y diagnósticos practicados advierten del empleo de instrumentos de evaluación escasamente fundamentados (Biscarri et al., 2006), en los que los requisitos normativos y administrativos parecen tener mayor incidencia que las necesidades derivadas del aprendizaje (Porto, 2006; Rizo, 2004; Scouller, 1996, 1998; Stobart, 2010).
- Otra de las grandes causas de inadecuación de las herramientas de evaluación empleadas refiere a la falta de validez y fiabilidad que se les atribuye como recursos de investigación (Stiggins, 2005a). Las dudas sobre la garantía técnica de los instrumentos utilizados lleva a cuestionar la capacidad de los sistemas evaluativos para clasificar el nivel de logro del alumnado con respecto a los fines previstos. Esa fue precisamente la conclusión que puso de manifiesto Burgess (2004, 2007) en su informe sobre la evaluación en las universidades del Reino Unido, al reconocer la incapacidad de los formatos sumativos finales y cuantitativos para describir las capacidades reales desarrolladas por las/os graduadas/os²².

h) Los sistemas tradicionales de evaluación en la universidad incurren en una excesiva simplificación del aprendizaje cuya valoración se pretende.

La evaluación tiende a limitarse a los aspectos estrictamente mensurables desde la convicción de que las calificaciones justas son aquellas matemáticamente justificables (Birenbaum, 1996; Bonsón & Benito, 2005). Desde esta perspectiva, la evaluación de los aprendizajes se centra en aquello que se considera fácil de valorar desde la lógica cuantitativa y objetivista, y se desprende de la consideración de aspectos más complejos que no se ajustan a la misma (R. McDonald, Boud, Francis, & Gonczi, 2000; Monescillo & Toscano, 2007). En consecuencia, se produce una excesiva simplificación de los aprendizajes valorados, a cambio de lograr cotas admisibles de fiabilidad (Sabirón & Arraiz, 2013; Stobart, 2010). “Los sistemas de evaluación más seguros, fiables y rigurosos a menudo van de la mano de aprendizajes superficiales y estériles y conducen a resultados de aprendizaje de

²² La conclusión llevó a determinar la necesidad de un nuevo sistema de calificación alternativo. Dicha aspiración se materializó años más tarde en la presentación del proyecto HEAR para el registro electrónico de los logros de aprendizaje de las/os universitarias a lo largo de su recorrido formativo (Burgess, 2012).

corta duración” (Gibbs & Simpson, 2009, p. 10). Sin embargo, la formación universitaria debería caracterizarse por la promoción de aprendizajes complejos y habilidades de alto nivel (Knight, 2005).

i) La evaluación se diseña y desarrolla como un proceso independiente de la enseñanza y el aprendizaje.

La evaluación del aprendizaje en la universidad tiene lugar como un proceso paralelo con tiempos y espacios diferentes (Ahumada, 2002, 2005). Brown (2003b) y Race (2003) proponen para la misma la metáfora del apéndice final, mientras que Rué (2007) se refiere a ella como un punto y aparte formativo.

Si para la evaluación la metáfora del punto y aparte no es la más indicada, la del punto y seguido mejora en algo las cosas, pero tampoco es una metáfora del todo feliz. En primer lugar, por evocar algo simple en vez de complejo. En segundo lugar, porque los puntos siempre son al final de la frase, cuando ésta ya ha sido construida. (p. 97).

Shepard (2000), Stiggins (2005a, 2005b) o Wiliam (2011), entre otros muchos, manifiestan que la evaluación representa un proceso consustancial a la enseñanza y el aprendizaje. Sin embargo, en la práctica se lleva a cabo desligada del aprendizaje, y desconectada de la enseñanza (O. Esteve, 2010), como si fuera una actividad con sentido en sí misma (Morales, 2008c; Rizo, 2004; D. Salinas, 2002).

j) La evaluación se muestra desalineada de los objetivos de aprendizaje pretendidos y las estrategias metodológicas desarrolladas para su consecución.

Frente al principio de alineación constructivista expuesto en el capítulo primero (Biggs, 2005; Biggs & Tang, 2011), en el contexto universitario español se acusan indicios de falta de correspondencia entre la evaluación practicada y los objetivos formativos y líneas de intervención metodológica establecidas para su logro (AQU, 2003; Mateo, 2006b, 2008). Este rasgo se relaciona directamente con la presunción de independencia desde la que se diseñan e implementan los dispositivos evaluativos, así como con la inadecuación de los instrumentos a los que recurren (Stiggins, 2002). Es especialmente acusado desde la perspectiva del alumnado, para quien las pruebas de evaluación no siempre guardan relación con el trabajo efectuado en el proceso, ni con los objetivos de aprendizaje difundidos (Otero, 2000;

Santos, 2009). No así para el profesorado, que aprecia niveles razonables de alineación en los sistemas de evaluación articulados (Gutiérrez et al., 2013; L. F. Martínez et al., 2012).

k) La evaluación del aprendizaje universitario no responde con contundencia a las exigencias de explicitación y transparencia.

La falta de transparencia y explicitación en los sistemas de evaluación en la universidad representa una queja tradicional ampliamente extendida en la literatura específica (Darby, 2007; Dochy et al., 2002; Hartley & Chesworth, 2000). Como en el caso de la continuidad de la evaluación y de la variedad de instrumentos, estudios recientes revelan un avance positivo en este sentido, probablemente por las exigencias de transparencia que lleva parejas el EEES (ENQA, 2009).

Tanto en el contexto nacional como en el internacional, la falta de información sobre los criterios y condiciones de la evaluación ha sido señalada como área de mejora en estudios desarrollados durante las tres últimas décadas (Burgess, 2004; M. Fernández, 1989; G. Rodríguez, 1998; Stiggins, 1999). Frente a esta situación inicial, análisis recientes de las programaciones docentes y de la normativa evaluativa han mostrado que los procesos de evaluación se desarrollan sobre condiciones debidamente informadas y explicitadas, en aras a garantizar la transparencia (Ibarra & Rodríguez, 2010; Martínez-Mínguez & Flores, 2014; Padilla et al., 2010; M. A. Sánchez, López, & Fernández, 2010). Así lo han concluido Gutiérrez et al. (2011) y L. F. Martínez et al. (2012) en los últimos años, aunque poniendo de manifiesto una desigual percepción de esta cuestión entre estudiantes y docentes.

l) La evaluación del aprendizaje en la universidad ofrece un feedback deficiente, insuficiente y poco frecuente.

El feedback producido en los procesos de evaluación del aprendizaje en la universidad resulta insuficiente tanto cualitativa como cuantitativamente (Gibbs & Simpson, 2009; Huxham, 2007). “La retroalimentación es un elemento crucial de la evaluación. Sin embargo, hay pruebas sustanciales de que el personal y los estudiantes están insatisfechos con él, de modo que un nuevo impulso para el cambio está creciendo” (OGG- ASKe, 2007, p. 2).

Aunque el profesorado señale que la evaluación debe ser pedagógica o favorecedora de la mejora, durante el proceso no se ofrece ningún tipo de feedback (Olmos & Rodríguez, 2010). De acuerdo con Stiggins (2005a) las prácticas

tradicionales de retroalimentación en la universidad no responden a las necesidades informativas de las audiencias interesadas, en particular, a las del alumnado. Así lo sugieren también Monescillo y Toscano (2007) y Castillo (2004), al afirmar que el feedback proporcionado desvirtúa su potencialidad didáctica por no ofrecer orientaciones para la mejora del aprendizaje.

En una revisión reciente sobre el tema Evans (2013) ha concluido que, en la última década, se ha logrado establecer la relevancia de la retroalimentación y una serie de condiciones para favorecer su calidad. Aun así, las dificultades para su puesta en práctica siguen siendo manifiestas según reflejan fuentes nacionales e internacionales. Entre las principales deficiencias que se atribuyen al feedback en el contexto universitario destacan las siguientes:

- *Insuficiencia.* Una de las deficiencias más básicas asociadas al feedback es su infrecuencia (Ferguson, 2011; Gibbs & Simpson, 2009; Gil-Flores, 2012; Race, 2003). En la medida en la que se hace de la evaluación una actividad puntual y final, el feedback adopta ese mismo carácter; se reduce a la mera comunicación de resultados y las aclaraciones sobre los mismos sólo se producen en caso de ser requeridas por el alumnado (Murphy, 2006).
- *Impuntualidad.* Los diagnósticos efectuados sobre los sistemas de evaluación del aprendizaje universitario revelan que, cuando se proporciona, el feedback tiende a llegar tarde, comprometiendo su utilización potencial con fines de mejora (Brown & Knight, 1994; Carless, 2006; J. Chappuis, 2012; Rué, 2009b).
- *Negatividad.* Se produce una acusada tendencia a centrar el feedback en los aspectos negativos detectados (Huxham, 2007; B. Salinas & Cotillas, 2007; Wootton, 2002). Los avances positivos en el proceso de aprendizaje tienden a obviarse y las pautas de proacción o sugerencias de mejora son infrecuentes. Ello provoca en el alumnado sensación de enjuiciamiento (Boud & Molloy, 2013).
- *Dificultades de interpretación.* Los estudios que se aproximan al punto de vista discente dan cuenta de dificultades en la interpretación de las orientaciones que desprende la retroalimentación (J. Chappuis, 2005, 2012; Walker, 2013). Éstas parecen asociadas a dos causas que pueden tener lugar simultáneamente. Por una parte, como indican Ferguson (2011) y Hounsell (2003) en ocasiones el feedback se reduce a comentarios vagos y difusos, insuficientemente elaborados como para resultar comprensibles. Por

la otra, la falta de familiarización con la propia disciplina también puede dificultar la decodificación del mensaje (Bloxxham, 2013; Carless, 2002, 2006; Knight, 2005; Price, Handley, Miller, & O'Donovan, 2010; Sadler, 2010). Independientemente de la causa, las dificultades reseñadas no sólo limitan la potencialidad didáctica del feedback, sino que además evocan sentimientos y sensaciones negativas.

Los estudiantes hablaron de sentimientos de ira y frustración cuando recibieron un feedback insuficientemente claro o desarrollado, y algunos hablaron explícitamente de su falta de confianza en los profesores y el proceso de evaluación. (Bryson, McDowell, McGugan, & Sanders, 2008, p. 2).

- *Desaprovechamiento.* Las dificultades de interpretación referidas hacen que parte de la retroinformación ofrecida por el profesorado no sea reconocida como tal por el alumnado y quede sin efecto (Jonsson, 2013; Orsmond & Merry, 2011; Sadler, 2013; Sambell, Gibson, & Montgomery, 2007). Esta circunstancia se produce principalmente cuando el profesorado recurre a cauces informales para aclarar los resultados de aprendizaje constatados y realizar recomendaciones de mejora (Higgins, Hartley, & Skelton, 2002; Price et al., 2010; Sambell, 2011, 2013).
- *Unilateralidad y énfasis en la calificación.* Por lo general, el feedback proporcionado en la enseñanza superior es calificativo y lo genera y administra el profesorado (Stiggins, 2005a). Tiene lugar por medio de un proceso comunicativo unilateral, en el que apenas se produce intervención discente (Cano, 2014; Nicol, 2010; Sadler, 2010; Sambell, 2011). Esta circunstancia incrementa la probabilidad de incidencia de las dos últimas dificultades.

m) La evaluación del aprendizaje en la universidad es homogénea o uniforme.

Deudora de una tradición formativa grupal y homogeneizadora, la evaluación del aprendizaje en la universidad es estereotipada. Uniforme y homogénea, no responde a la diversidad natural existente entre las/os estudiantes, ni tampoco la considera (Ahumada, 2002, 2005; Coll et al., 2000; Santos, 1999; Trillo, 2005). Como aborda el próximo capítulo, los referentes normativos para la praxis evaluativa

refuerzan la objetividad y uniformidad de las condiciones evaluativas como rasgo deseable (Ibarra & Rodríguez, 2010; Padilla et al., 2010).

n) La evaluación del aprendizaje que se lleva a cabo en la universidad, favorece por parte de las/os estudiantes la adopción de enfoques estratégicos y superficiales de aprendizaje.

El poder estratégico de la evaluación en el aprendizaje puede tornarse contraproducente cuando ésta no responde a los objetivos formativos previstos (Brown, 2003b). La simplificación a la que se somete el aprendizaje para hacerlo directamente evaluable por medio de instrumentos cuantitativos de fácil manejo y administración potencia entre el alumnado el empleo de las funciones intelectuales menos ricas (Biggs, 2005; Santos, 1999). Las demandas fácilmente cuantificables rara vez refieren a aprendizajes complejos y profundos. En este contexto, incrementa la probabilidad de que se produzcan aproximaciones superfluas y estratégicas (I. Álvarez, 2008a; Gibbs & Simpson, 2009; Hernández, García, & Maquilón, 2001; Monereo & Castelló, 2009b), caracterizadas por el análisis de las exigencias evaluativas y la orientación exclusiva de los esfuerzos a la satisfacción de las mismas (Entwistle, 2000; Marton & Säljö, 1984/2005).

o) La evaluación del aprendizaje universitario es fundamentalmente convergente.

A consecuencia de la tendencia reproductiva y declarativa de la que dan cuenta los instrumentos empleados, la evaluación en la universidad resulta fundamentalmente convergente (Biggs, 2005). A diferencia de lo que sucede en el ámbito profesional real, por lo general se desarrollan tareas evaluativas simples, que admiten una única respuesta correcta (I. Álvarez & Gómez, 2009; Dochy et al., 2002). Las tareas divergentes dan lugar a procesos evaluativos más complejos, menos objetivos y más polémicos, por lo que no son frecuentes. Esta tendencia a la convergencia, unida a la homogeneización, conduce a sistemas de evaluación en los que las/os estudiantes replican la misma respuesta, al mismo nivel y en igualdad de condiciones (Bilbao, 2010).

p) La evaluación en la universidad se realiza sobre la base de un aprendizaje atomizado o parcelado desde la perspectiva disciplinar.

Además de referirse básicamente a aprendizajes declarativos, las evaluaciones en la universidad se construyen sobre la base de un conocimiento parcelado, no de forma integrada ni globalizada. “Mientras las demandas actuales a los procesos, tanto de enseñanza como de aprendizaje, apuntan a que estén impregnados de una

perspectiva globalizadora los sistemas siguen desgastándose en enfoques parciales que generan sólo conocimientos estancos” (Ahumada, 2005, p. 21). Analizando la cultura escolar, J. Vera y Esteve (2001) coinciden en describir una praxis evaluativa basada en exámenes que exigen el dominio de contenidos teóricos presentados de forma atomizada. No son frecuentes las demandas que exigen una aplicación integrada de las distintas disciplinas. Sin embargo, la suma de varias partes no necesariamente garantiza la consecución de los aprendizajes previstos (Jessop et al., 2014; Porto, 2006; Rizo, 2004; Sabirón & Arraiz, 2013), ni su reproducción asegura la capacidad de aplicación.

q) *La evaluación del aprendizaje en la universidad se produce de forma descontextualizada.*

La evaluación del aprendizaje en los niveles universitarios se lleva a cabo en situaciones y condiciones alejadas de la práctica real (Gulikers, Bastiaens, & Kirschner, 2004; MacLellan, 2001; Murphy, 2006).

Si revisamos la mayoría de las pruebas y exámenes que se realizan habitualmente en las universidades podremos comprobar que la mayoría de ellas no responden a esta idea de una evaluación auténtica, sino que más bien son pruebas de memorística o de resolución de problemas más o menos abstractos y, en muchas ocasiones, claramente alejados de la realidad o de la práctica profesional para la que supuestamente se está formando. (A. Pérez et al., 2009, p. 33).

r) *La evaluación del aprendizaje de las/os estudiantes en la universidad tiene una dudosa capacidad de predicción de la actuación profesional en contextos reales.*

Unido a lo anterior, las condiciones en las que se llevan a cabo las evaluaciones del aprendizaje universitario no predicen la actuación en el futuro desempeño profesional (Gibbs & Simpson, 2009; Monereo & Castelló, 2009b). Stobart (2010) y Race (2003) advierten que, en ocasiones, el éxito en la evaluación depende de la técnica o pericia que haya desarrollado la/el estudiante ante la propia situación evaluativa, no del aprendizaje logrado.

Arribas et al. (2010) mostraron un calado aún mayor de esta limitación al constatar que, desde el punto de vista del alumnado egresado, la formación básica de

maestras/os concluye sin que se hayan desarrollado algunas de las competencias básicas para el ejercicio profesional.

s) La evaluación del aprendizaje en la universidad responde más a las condiciones prácticas y a las exigencias normativas y administrativas que a las necesidades de las/os estudiantes.

Variables como el número de estudiantes, los calendarios prefijados de evaluación o las condiciones detalladas en el marco regulador de referencia (formatos, convocatorias, excepciones, etc.) representan una poderosa influencia en las decisiones del profesorado sobre la evaluación del aprendizaje (ARC, 2009d; M. L. Pérez et al., 2000; Porto, 2010; Zaragoza, Luis-Pascual, & Manrique, 2009). El diseño de los dispositivos evaluativos a veces depende más de este tipo de circunstancias prácticas, administrativas o normativas, que de las necesidades que se desprenden del propio proceso de enseñanza y aprendizaje (Batallosa, 2004; Navarro, Santos, Buscà, Martínez, & Martínez, 2010). Así lo expresan Sabirón y Arraiz (2013):

Es difícil determinar el resultado efectivo de un aprendizaje, pero relativamente fácil su calificación tras una evaluación programada. Desde la enseñanza infantil a la universitaria, se acumulan, a golpe de reglamento, procedimientos y formas que pretenden someter la realidad, complicada y compleja, a la previsión programada. El discurso ya no es exclusivo de la pedagogía del docente, pues pertenece más bien en nuestros días al sistema y la institución. La normativa y el procedimiento regulan una programación apropiada por la burocracia de la administración, gestionada quizá por un docente, pero cada vez más ajena al profesor, evitándole incluso la posibilidad de caer en sus propias contradicciones (que también podían resultar educativas). (p. 20).

t) En la universidad apenas se llevan a cabo procesos de metaevaluación.

En los contextos formativos universitarios no es frecuente someter a evaluación la propia evaluación, es decir, no se practica la metaevaluación (B. McDonald, 2010; Monescillo & Toscano, 2007; Santos & Moreno, 2004). Los programas no son

sometidos a revisión sistemática y así coinciden en afirmarlo estudiantes, docentes y alumnado egresado (Arribas et al., 2010).

Santos (1999) asocia esta deficiencia a la reproducción sucesiva de los sistemas evaluativos aplicados en la universidad año tras año. De acuerdo con S. Rodríguez (2005), otra de sus consecuencias indirectas es el habitual desconocimiento de las buenas prácticas en el seno de la institución, la titulación o incluso dentro de un mismo departamento. Este último aspecto, conecta con la siguiente deficiencia.

u) Los sistemas de evaluación del aprendizaje universitario ofrecen señales evidentes de descoordinación profesional al nivel de titulación.

Cebrián (2008) habla de *autismo pedagógico* para referirse a la tradición formativa que permite la coexistencia de prácticas diferentes que se ignoran mutuamente dentro del sistema educativo o en una misma institución. Así se refleja en la evaluación del aprendizaje que tiene lugar en las universidades españolas, donde no existe una visión compartida sobre su calidad y alcance (Buendía & López, 1999; Cano et al., 2010; Porto, 2008). En consecuencia, las/os estudiantes perciben descoordinación y grandes diferencias entre las exigencias evaluativas que se les plantean en los diversos módulos o asignaturas (Arribas et al., 2010; Trillo & Porto, 2002).

La variedad en los sistemas evaluativos representa una manifestación natural de la diversidad de aprendizajes a valorar. Esta deficiencia hace referencia a la descoordinación en aspectos más generales como el tipo de demandas que se priorizan, el volumen y la distribución de trabajo o el rigor con el que son aplicadas las condiciones definidas.

v) La evaluación del aprendizaje en la universidad no complace a docentes, a discentes, ni a audiencias externas.

El cúmulo de disfuncionalidades y limitaciones descrito se traduce en una práctica evaluativa que no complace a estudiantes, a docentes ni a ninguna otra de las audiencias involucradas (evaluadores externos, familias, empleadores, etc.) (Jorba & Sanmartí, 2000; M. P. Romero, 2004).

Aunque las manifestaciones del alumnado tengan mayor sonoridad, la descripción de vivencias y predisposiciones negativas con respecto a la evaluación se produce tanto en docentes, como en estudiantes (I. González, 2003; Stiggins, 1999; Trillo, 2005).

1.2 Resumen de las principales deficiencias.

Las carencias y limitaciones descritas a lo largo de las páginas anteriores hacen referencia a una misma realidad y mantienen entre sí una estrecha relación de interdependencia difícil de precisar. La tabla 18 sintetiza en cinco descriptores generales las principales disfuncionalidades o limitaciones reportadas, pero el mapa de relaciones e influencias entre los rasgos resumidos puede adoptar formas e interpretaciones diversas. La síntesis resultante reúne bajo un mismo descriptor aquellos rasgos sobre los que presumimos una mayor proximidad e interacción, admitido que caben otras agrupaciones.

Tabla 18

Descriptores generales de los sistemas tradicionales de evaluación del aprendizaje en la universidad

La típica evaluación del aprendizaje en la universidad es principalmente...	
A	<i>Sumativa y final, realizada unilateralmente por el profesorado sobre productos individuales del aprendizaje del alumnado. Aunque se aprecia la proliferación de formatos de evaluación continua, se mantiene la sobrerrepresentación de la finalidad acreditativa y del profesorado como principal agente evaluador (único, la mayor parte de las veces).</i>
B	<i>Independiente del proceso de enseñanza-aprendizaje y desalineada, escasamente planificada y basada en decisiones poco fundamentadas, que responden más a condiciones prácticas, normativas y administrativas, que a las necesidades pedagógicas derivadas del aprendizaje.</i>
C	<i>Homogénea y convergente, basada en la reproducción de contenidos conceptuales fragmentados en situaciones de evaluación descontextualizadas, a las que se puede responder mediante aproximaciones estratégicas o superficiales al aprendizaje.</i>
D	<i>Monometódica, basada en la realización de exámenes o pruebas, a menudo inadecuadas, que simplifican el aprendizaje y carecen de capacidad predictiva con respecto a su transferencia práctica. Aunque se aprecia la proliferación de nuevas herramientas evaluativas, el examen sigue siendo el principal por frecuencia e incidencia en la calificación.</i>
E	<i>Incoherente, que no complace a ninguna de las audiencias más directamente involucradas, que genera incomodidades y sentimientos negativos en todas ellas y una cierta tendencia de evitación o negación general (apenas se hace metaevaluación, coordinación o intercambio de experiencias).</i>

En términos generales, la revisión efectuada nos lleva a concluir que el sistema universitario de evaluación del aprendizaje es disfuncional, calificativo que denota un desarreglo en el funcionamiento de algo o en la función que le corresponde (Real Academia Española, 2014). Así se constata al explorar las dos principales funcionalidades atribuidas a la evaluación en el anterior capítulo: la función acreditativa, que tiene lugar con indicios que ponen en entredicho su validez y fiabilidad (incapacidad predictiva, imprecisión, valoración de aspectos parciales y poco representativos del aprendizaje, etc.) y la función pedagógica, escasamente explotada e infrarrepresentada con respecto a la anterior. Según

Batalloso (2004), “sabemos que las rutinas de la evaluación están afectadas de una patología que las desnaturaliza, patología con la que la evaluación pierde su sentido de comprensión y mejora de las acciones educativas” (p. 50).

Cuanto más disfuncional resulte un sistema de evaluación del aprendizaje, menor será su potencialidad didáctica y mayor será la probabilidad de provocar efectos adversos. Especialmente evidentes en las audiencias más implicadas en el proceso (alumnado y profesorado), dichos efectos constituyen el objeto central del siguiente apartado.

2. Efectos de los sistemas tradicionales de evaluación del aprendizaje en el alumnado y el profesorado universitario

Las consecuencias de una mala praxis evaluativa son nefastas, y así lo advierten entre otros Stobart (2010), McDowell y Sambel (2003) o Stiggins (1999, 2005b). Su sombra es alargada (D. Salinas, 2002) y repercute en el alumnado y en otras audiencias interesadas como el profesorado, la propia institución formativa, las familias de las/os estudiantes, las asociaciones profesionales, las/os empleadoras/es potenciales y, en definitiva, la sociedad, que precisa de ciudadanas/os y profesionales competentes.

Abarcar este amplio abanico de consecuencias y repercusiones excede los límites de la presente investigación. En consecuencia, los próximos apartados restringen la atención a los colectivos con mayor vinculación: el alumnado y el profesorado universitario, las mismas fuentes que serán recurridas en el estudio empírico.

2.1 Efectos de los sistemas disfuncionales de evaluación del aprendizaje en el alumnado universitario.

La evaluación da lugar a decisiones que afectan al futuro y al desarrollo educativo del alumnado (Bonsón & Benito, 2005; JCSEE, 2007). Desde esa premisa Boud (1995) afirma que las/os estudiantes pueden escapar a una mala enseñanza, pero nunca evitar una mala evaluación. La investigación revela que el empleo de sistemas de evaluación

continua, con una marcada orientación pedagógica, favorece en el alumnado un mejor rendimiento académico (Arribas, 2012; Castejón, López, Julián & Zaragoza, 2012; Fraile, López, Castejón, & Romero, 2013; Holmes, 2015; R. Romero, Fraile, López, & Castejón, 2014). Sin embargo, tales condiciones no parecen extendidas en la enseñanza universitaria, según la situación resumida en el anterior apartado.

La literatura específica describe efectos de naturaleza diversa al explorar la incidencia de las deficiencias evaluativas en el alumnado universitario. Realizada una revisión preliminar de los mismos, consideramos que los más relevantes quedan recogidos en los cuatro núcleos propuestos a continuación:

a) *El efecto de las calificaciones y el expediente académico como mecanismo de selección académico y profesional.*

El sistema académico y profesional en España tiene un marcado carácter meritocrático, de modo que las calificaciones y el expediente académico representan un importante mecanismo de selección. Aunque desde la pasada década hayan surgido nuevas demandas y filtros selectivos (Gairín & Darder, 1994), la participación en proyectos y experiencias formativas nacionales e internacionales o el acceso a becas sigue dependiendo, en gran medida, de los logros académicos acreditados. Este efecto discriminador se fundamenta en la presunción de que las calificaciones evidencian el desarrollo de un determinado paquete de aprendizajes o competencias. Sin embargo, como señala Flemming (2003), esta asunción se construye suponiendo que la evaluación del aprendizaje discente es válida, relevante y fiable, cuando los síntomas enumerados en el apartado anterior sugieren más bien lo contrario.

b) *La connotación o vertiente negativa de la evaluación.*

Santos (2009) da cuenta de un componente amargo y doloroso en los relatos del alumnado universitario sobre la evaluación, sujeto a dos posibles explicaciones: la memoria selectiva, que amplifica el efecto de las vivencias negativas o la incidencia de errores e incongruencias reales de la evaluación experimentada.

Entre las principales manifestaciones reportadas, se encuentran las siguientes:

- Insatisfacción y frustración académica, manifestada en sentimientos de desaliento, frustración y fracaso (M. Fernández, 1989; Flórez, 1999; V. M. López, 2004). Estrechamente apegada a la dimensión emocional de la evaluación (Boud, 1995; Falchikov & Boud, 2007), la vivencia reiterada de este tipo de sensaciones podría conducir a abandono en casos extremos.

- Unido a lo anterior, la destrucción del autoconcepto del alumnado es otra de las secuelas que puede dejar tras de sí un sistema de evaluación del aprendizaje indebidamente calibrado (Arter, 2009; Assessment Reform Group-ARG, 2002; Race, 2003). Pueden conducir a este término exigencias evaluativas excesivas (Brown, Race, & Smith, 1996; Gibbs & Simpson, 2009; Monereo & Castelló, 2009b; Nicol, 2007,2009) o la acumulación de evaluaciones realizadas desde la perspectiva negativa, donde se evidencian más los errores y fracasos que los logros y avances (Ferguson, 2011; Wootton, 2002).
- La evaluación genera en el alumnado desconfianza y un cierto afecto de agresividad (Bryson et al., 2008). Empleada por el profesorado como instrumento de control (Batalloso, 2004; Monescillo & Toscano, 2007), el sistema de evaluación universitario descrito sitúa a las/os estudiantes en una posición de sumisión y desventaja desde la que se explican tales sentimientos.
- Muy frecuentemente el alumnado refiere estrés, presión o nerviosismo al describir su experiencia evaluativa en la universidad (L. F. Martínez et al., 2012; Ricoy & Fernández-Rodríguez, 2013).
- La vivencia del éxito y del fracaso académico en función de la calificación, es otra de las experiencias evaluativas con repercusión en las/os estudiantes (Litwin, 1998). Respaldado desde dentro y fuera del sistema educativo, este énfasis en la calificación lleva al alumnado a saberse *etiquetado* desde edades muy tempranas (Santos, 2009; Stiggins, 2007).
- La meritocracia individualista es otra de las secuelas que deja en el alumnado una práctica disfuncional de la evaluación del aprendizaje. Según V. M. López (2004) o Sanmartí y Jorba (1995), ésta alimenta la competitividad por encima de la colaboración y la cooperación entre estudiantes.

c) Los efectos de la evaluación en el aprendizaje explícito.

En la presente investigación el aprendizaje se concibe como el rendimiento evidenciado en el desarrollo de las competencias previstas en el perfil profesional definido para la titulación. Interpretado desde la óptica constructivista, se trata de un proceso complejo originado a partir de la experiencia, que da lugar a cambios permanentes en el individuo (Etxague, 2004). Sin embargo, las/os estudiantes

universitarias/os tardan poco tiempo en olvidar lo tratado en las diversas materias (Santos, 2009) y a menudo mantienen concepciones erróneas sobre las mismas, incluso habiendo aprobado (Schommer et al., 2012).

Esta apreciación sugiere que, a pesar de ser empleada con fines acreditativos, la evaluación no garantiza realmente la consecución del aprendizaje necesario para el desempeño profesional. Orientada principalmente al conocimiento teórico o declarativo, el alumnado logra superar las exigencias evaluativas con paquetes de aprendizajes memorísticos y ello no garantiza su capacidad de transferencia y puesta en práctica (Biggs, 2005; Zabalza, 2006).

d) *Los aprendizajes implícitos sobre evaluación.*

Los aprendizajes que las/os estudiantes realizan en la universidad no sólo tienen que ver con las competencias generales y específicas de la titulación, sino que afectan también a sus concepciones sobre la naturaleza del propio aprendizaje, la enseñanza, la profesionalidad o el valor de la formación. Por los propósitos de la presente investigación, resultan de particular interés los aprendizajes que construye el alumnado con respecto a la evaluación.

La experimentación de patrones evaluativos recurrentes lleva al alumnado a realizar aprendizajes sobre los propios sistemas de evaluación.

Aprenden que aprobar los exámenes es mucho más importante que saber y tener curiosidad intelectual o motivación por el conocimiento; aprenden que obedecer al profesor y decirle lo que éste espera oír es mucho más rentable que expresar sus propias opiniones; aprenden que si se hacen demasiado visibles o protestan en exceso estarán sujetos a un mayor control calificadorio; aprenden que si tienen suerte pueden aprobar sin esfuerzo; aprenden, en suma, un conjunto de trucos que en nada tienen que ver ni con la cultura ni con la educación ni con el placer de saber y descubrir. (Batalloso, 2004, pp. 46-47).

La tabla 19 presenta el inventario de aprendizajes probables que puede dejar tras de sí la exposición continuada a los sistemas disfuncionales de evaluación descritos en el apartado anterior:

Tabla 19

Efectos de los principales rasgos de disfuncionalidad evaluativa en el alumnado universitario: aprendizajes implícitos sobre la evaluación

La típica evaluación del aprendizaje en la universidad es principalmente...	
A	<i>Sumativa y final, realizada unilateralmente por el profesorado sobre productos individuales del aprendizaje del alumnado. En consecuencia, las/os estudiantes...</i> • Aprenden que lo importante es el producto, y no tanto el proceso. • Aprenden que la evaluación de trabajos realizados en equipo no es justa y, en consecuencia, aprenden a manejarse en esa situación poniendo en juego estrategias diferentes, no todos admisibles desde la perspectiva ética y pedagógica. • Aprenden que los errores son algo negativo y que, en consecuencia hay que evitarlos u ocultarlos si se producen. • Aprenden que a veces es más sencillo aprobar un examen final que someterse a las exigencias de un dispositivo de evaluación continua. • Aprenden que importa más la calificación que el aprendizaje: no aprobar parece más grave, que haber aprobado sin aprender nada. • Aprenden que la/el docente califica y el alumnado es calificado. • Aprenden a mantener en silencio sus desacuerdos y recelos con respecto a los procesos evaluativos, por temor a venganzas o represalias.
B	<i>Independiente del proceso de enseñanza-aprendizaje y desalineada, escasamente planificada, y basada en decisiones poco fundamentadas que responden más a condiciones prácticas, normativas y administrativas, que a las necesidades pedagógicas derivadas del aprendizaje. En consecuencia, las/os estudiantes...</i> • Aprenden que el examen no siempre se corresponde con los objetivos expuestos y el trabajo efectuado en clase. • Aprenden que una cosa es la enseñanza y el aprendizaje, y otra diferente la evaluación (momentos y espacios diferentes, distinto tipo de demandas, etc.). • Aprenden que el profesorado debe ceñir la evaluación a una normativa que regula plazos y condiciones y que repercute en el tipo de demandas que se plantean. • Aprenden que la elección de un sistema u otro de evaluación tiene mucho que ver con el número de estudiantes o la carga de trabajo que supone para el profesorado. • Aprenden que la elección de un sistema u otro de evaluación no tiene que ver con sus necesidades formativas.
C	<i>Homogénea y convergente, basada en la reproducción de contenidos conceptuales fragmentados, en situaciones de evaluación descontextualizadas, a las que se puede responder mediante aproximaciones estratégicas y superficiales al aprendizaje. En consecuencia, las/os estudiantes...</i> • Aprenden que las tareas y demandas evaluativas están sujetas a una única alternativa de solución, aquella dictaminada por el agente evaluador. • Aprenden que es más seguro reproducir aquello que espera el profesorado, que ejercitar la capacidad crítica y creativa. • Aprenden a identificar las preferencias de las/os distintas/os profesoras/es con respecto a la evaluación. • Aprenden que es posible superar un examen realizando un trabajo de retención memorística concentrada en los días previos. • Aprenden que no haber realizado ningún esfuerzo durante el ciclo de enseñanza-aprendizaje y aprobar el examen pueden ser circunstancias compatibles.

La típica evaluación del aprendizaje en la universidad es principalmente...

- Aprenden que las condiciones de la evaluación rara vez son negociables o adaptables a las necesidades particulares del alumnado.
- Aprenden que las adecuaciones en las condiciones de la evaluación son un reflejo de agravios comparativos o conductas arbitrarias por parte del agente evaluador.

D

*Monometódica, basada en la realización de exámenes o pruebas a menudo inadecuadas, que simplifican el aprendizaje y carecen de capacidad predictiva con respecto a su transferencia a la práctica. **En consecuencia, las/os estudiantes...***

- Aprenden que, con independencia del trabajo efectuado durante el curso, el examen representa la hora de la verdad y el más importante de los componentes calificativos.
- Aprenden que, ante todo y sobre todo, lo más importante es superar el examen, aunque ello suponga incurrir en conductas poco éticas y/o saludables.
- Aprenden que el aprendizaje se puede descomponer en una artificial suma de elementos que llevan a una puntuación final.
- Aprenden que los procesos de corrección no siempre son precisos y que pueden producirse desviaciones a favor o en contra, en función de una serie de factores (preferencias docentes sobre presentación, extensión y estilo de redacción, expectativas previas generadas en la dinámica del aula, etc.).
- Aprenden que, una vez superado el examen, pueden despojarse del contenido retenido para liberar espacio en su disco duro. Indirectamente, aprenden que en el ejercicio profesional teoría y práctica son independientes.

E

*Incoherente, que no complace a ninguna de las audiencias más directamente involucradas, que genera incomodidades y sentimientos negativos en todas ellas y una cierta tendencia de evitación o negación general. **En consecuencia, las/os estudiantes...***

- Aprenden a desconfiar de la evaluación, a recelar o sospechar sobre el rigor de la misma.
- Aprenden que en evaluación una cosa es lo que se dice y otra muy diferente lo que se hace.
- Aprenden que las limitaciones de la evaluación ni se admiten, ni se solucionan.
- Aprenden que la evaluación es un mal necesario en la vida estudiantil, una prueba que hay que superar, no una oportunidad o recurso para el aprendizaje.

Nota: Elaboración propia a partir de las aportaciones generales de R. McDonald et al. (2000), Batalloso (2004), Santos (1999) y Ricoy y Fernández-Rodríguez (2013).

Todos estos aprendizajes van configurando una determinada concepción o forma de entender la evaluación. Estas representaciones o concepciones evaluativas van a ocupar un lugar destacado a la hora de analizar las causas de disfuncionalidad de los sistemas evaluativos, en especial, en la formación inicial de maestras/os.

2.2 Efectos de los sistemas disfuncionales de evaluación del aprendizaje en el profesorado universitario.

Aunque las referencias son mucho menos frecuentes en la literatura, los efectos de los sistemas disfuncionales de evaluación del aprendizaje también afectan al profesorado universitario. La síntesis de las aportaciones recopiladas en este sentido nos lleva a clasificar cuatro bloques interconectados de consecuencias:

a) La vertiente negativa de la evaluación para el profesorado.

La evaluación genera en el profesorado inquietudes y preocupaciones asociadas a su dimensión ética (justicia, gravedad de los efectos de las calificaciones asignadas, etc.) y, con ellas, una fuente potencial de desequilibrio emocional (Trillo, 2005). Representa una tarea ineludible (Casanova, 1992), molesta y tediosa (Flórez, 1999; Serrano, 2010); sin embargo ocupa una parte importante de la actividad docente (Atkinson & Lim, 2013; Mertler, 2009; Stiggins, 2005a; Stiggins & Conklin, 1988).

b) Los efectos de la evaluación en la efectividad de la enseñanza.

Este trabajo se sitúa en el marco de la evaluación para el aprendizaje, caracterizado por confiar en el valor y poder de la actividad evaluativa como recurso para estimular estratégicamente en el alumnado el logro de los fines previstos. En contrapartida, “si las evaluaciones en el aula son de mala calidad, entonces la instrucción no puede ser efectiva” (Stiggins, 1999, p. 26). Desde esta perspectiva, los sistemas disfuncionales de evaluación restan potencialidad didáctica al proceso general de enseñanza. Dependiendo de cómo se emplee, la evaluación puede compensar o descompensar aún más otras limitaciones asociadas a la docencia (Morales, 2008c).

c) Disminución de las creencias de autoeficacia docente.

“La autoeficacia es una creencia o conjunto de creencias que tiene el propio sujeto con respecto a si será capaz o no de llevar a cabo algo y conseguir un nivel adecuado u óptimo de realización” (Torre, 2008, p. 62). Partiendo de esa base, consideramos que los procesos atributivos que realice el profesorado para explicar las disfuncionalidades de la evaluación pueden afectar a sus creencias de autoeficacia docente, dependiendo de las causas a las que apunten. Cuando éstas estén ligadas a factores fuera de su ámbito de incumbencia profesional (falta de recursos, normativa, etc.) no parece probable que así suceda; sin embargo, si las disfuncionalidades se atribuyen a factores o causas internas, como la confusión

(Arénaga, 2013; V. M. López et al., 2007; Rué, 2008) o la falta de preparación (Palacios & López, 2013), la autoimagen profesional podría verse dañada.

d) Los efectos de la evaluación en el desarrollo profesional docente y la profesionalidad: conformismo y aprendizajes implícitos sobre evaluación.

La amplia vigencia de los sistemas disfuncionales de evaluación de un lado, y la falta de cuestionamiento de las prácticas evaluativas del otro, provocan en el profesorado aprendizajes implícitos sobre el fenómeno que nos ocupa. Son aprendizajes que tienen lugar en el ámbito de lo profesional y que pueden llegar a interferir en las dimensiones ética, técnica y/o didáctico-pedagógica de la función evaluativa. Como en el caso del alumnado, la tabla 20 recoge alguno de los ejemplos más significativos que pueden producirse en este sentido:

Tabla 20

Efectos de los principales rasgos de disfuncionalidad evaluativa en el profesorado universitario: aprendizajes implícitos sobre la evaluación que generan desprofesionalización

La típica evaluación del aprendizaje en la universidad es principalmente...	
A	<i>Sumativa y final, realizada unilateralmente por el profesorado sobre productos individuales del aprendizaje del alumnado. En consecuencia, las/os docentes...</i> • Aprenden que evaluar es plantear tareas y/o pruebas (exámenes) al alumnado, corregirlas y trasladarlas a la calificación. • Aprenden que la evaluación de tareas realizadas en equipo a menudo acarrea conflictos, aunque alivia la carga de trabajo. • Aprenden que la evaluación exclusivamente final es difícilmente justificable y que la evaluación continua incrementa la carga de trabajo. En consecuencia... • Aprenden a diseñar sistemas de evaluación continua basados en demandas concretas y de fácil cuantificación, que den lugar a una calificación final clara. • Aprenden que lo que interesa al alumnado es la calificación y que eso les confiere poder. • Aprenden que lo que interesa a las/os estudiantes y al resto de las audiencias involucradas (institución formativa, familias, empleadores, etc.) es la calificación, no el aprendizaje ni el proceso formativo desarrollado. • Aprenden que la evaluación sólo repara en el alumnado y su trabajo.
B	<i>Independiente del proceso de enseñanza-aprendizaje y desalineada, escasamente planificada, y basada en decisiones poco fundamentadas que responden más a condiciones prácticas, normativas y administrativas, que a las necesidades pedagógicas derivadas del aprendizaje. En consecuencia, las/os docentes...</i> • Aprenden que una cosa es el trabajo de clase y la enseñanza y otra la evaluación. • Aprenden que la evaluación es una decisión que se adopta en el transcurso del proceso formativo o al finalizar el mismo, cuando se vaya a realizar. • Aprenden que determinados sistemas o formatos de evaluación generan al docente una sobrecarga difícil de sobrellevar. • Aprenden que es más conveniente emplear formatos de evaluación económicos, fáciles de administrar y corregir.

La típica evaluación del aprendizaje en la universidad es principalmente...

- Aprenden que compartir con el alumnado la información relacionada con las condiciones y criterios de evaluación limita su capacidad de actuación y reduce su poder.

C

*Homogénea y convergente, basada en la reproducción de contenidos conceptuales fragmentados, en situaciones de evaluación descontextualizadas, a las que se puede responder mediante aproximaciones estratégicas y superficiales al aprendizaje. **En consecuencia, las/os docentes...***

- Aprenden que proponer preguntas o situaciones en las que la respuesta sea única y clara simplifica la corrección y reduce la probabilidad de conflictos y desacuerdos.
- Aprenden que tareas evaluativas complejas (proyectos, casos, problemas, etc.) requieren mayor trabajo de seguimiento y corrección y que son fuente de conflictos.
- Aprenden que determinadas/os estudiantes logran superar las pruebas/exámenes a pesar de no haber efectuado un trabajo adecuado ni suficiente a lo largo del curso.
- Aprenden que las condiciones bajo las que se lleva a cabo la evaluación deben ser iguales para todas/os las/os estudiantes a fin de evitar conflictos.
- Aprenden que la atención a la diversidad desde la evaluación es un problema.
- Aprenden que la atención a la diversidad desde la evaluación es algo extraordinario sujeto a excepcionalidad.

D

*Monometódica, basada en la realización de exámenes o pruebas a menudo inadecuadas, que simplifican el aprendizaje y carecen de capacidad predictiva con respecto a su transferencia a la práctica. **En consecuencia, las/os docentes...***

- Aprenden que el examen representa la hora de la verdad, aunque su resultado no siempre refleje el trabajo efectuado durante el curso.
- Aprenden que los exámenes son la técnica evaluativa que mejor responde a los requisitos de objetividad, justicia y cuantificación.
- Aprenden a identificar los aspectos más fácilmente cuantificables para diseñar pruebas de corrección sencilla, que eviten la realización de valoraciones subjetivas y los conflictos consiguientes.
- Aprenden que la relación de las/os estudiantes con el contenido evaluado finaliza una vez se supera el examen.

E

*Incoherente, que no complace a ninguna de las audiencias más directamente involucradas, que genera incomodidades y sentimientos negativos en todas ellas y una cierta tendencia de evitación o negación general (no se hace metaevaluación, ni investigación, ni intercambio de experiencias). **En consecuencia, las/os docentes...***

- Aprenden que la evaluación que realizan del aprendizaje de las/os estudiantes forma parte de su ámbito profesional privado.
- Aprenden que habitualmente en evaluación una cosa es lo que se escribe y otra diferente lo que se hace.
- Aprenden que determinadas prácticas evaluativas se siguen manteniendo en el tiempo, a pesar de las limitaciones aparentes y reconocidas.
- Aprenden que la evaluación es una obligación sujeta a la docencia, una tarea administrativa de selección y acreditación que hay que acometer (no un recurso para la enseñanza y el aprendizaje).

Estos aprendizajes y experiencias condicionan las concepciones o representaciones de las/os docentes sobre la evaluación y su forma de emplearla en el proceso de enseñanza-aprendizaje.

3. Factores asociados a las principales deficiencias en la evaluación del aprendizaje universitario y su persistencia

Partiendo de la descripción general del modelo tradicional de evaluación universitario y de sus principales efectos en docentes y discentes, este apartado tiene por objeto analizar las causas y factores que explican las disfuncionalidades señaladas, así como su persistencia a lo largo del tiempo.

La literatura específica hace referencia a variables numerosas y de naturaleza diversa para explicar el origen y mantenimiento de este cuadro de deficiencias. Para su análisis proponemos clasificarlas en dos bloques:

- *Factores o condiciones estructurales.* Se incluyen en este primer bloque todos aquellos factores de orden legal, institucional, social u organizativo que condicionan los procesos evaluativos que tienen lugar en la universidad y el origen y mantenimiento de las deficiencias observadas en los mismos. Representan condiciones sobre las que el profesorado rara vez tiene un control directo.
- *Factores o condiciones profundas.* El calificativo profundo se emplea para hacer referencia a todas aquellas variables que, sin ser de orden estructural, explican la incidencia de los síntomas de disfuncionalidad en los sistemas evaluativos y su mantenimiento a lo largo del tiempo. Representan constructos más complejos que las condiciones estructurales.

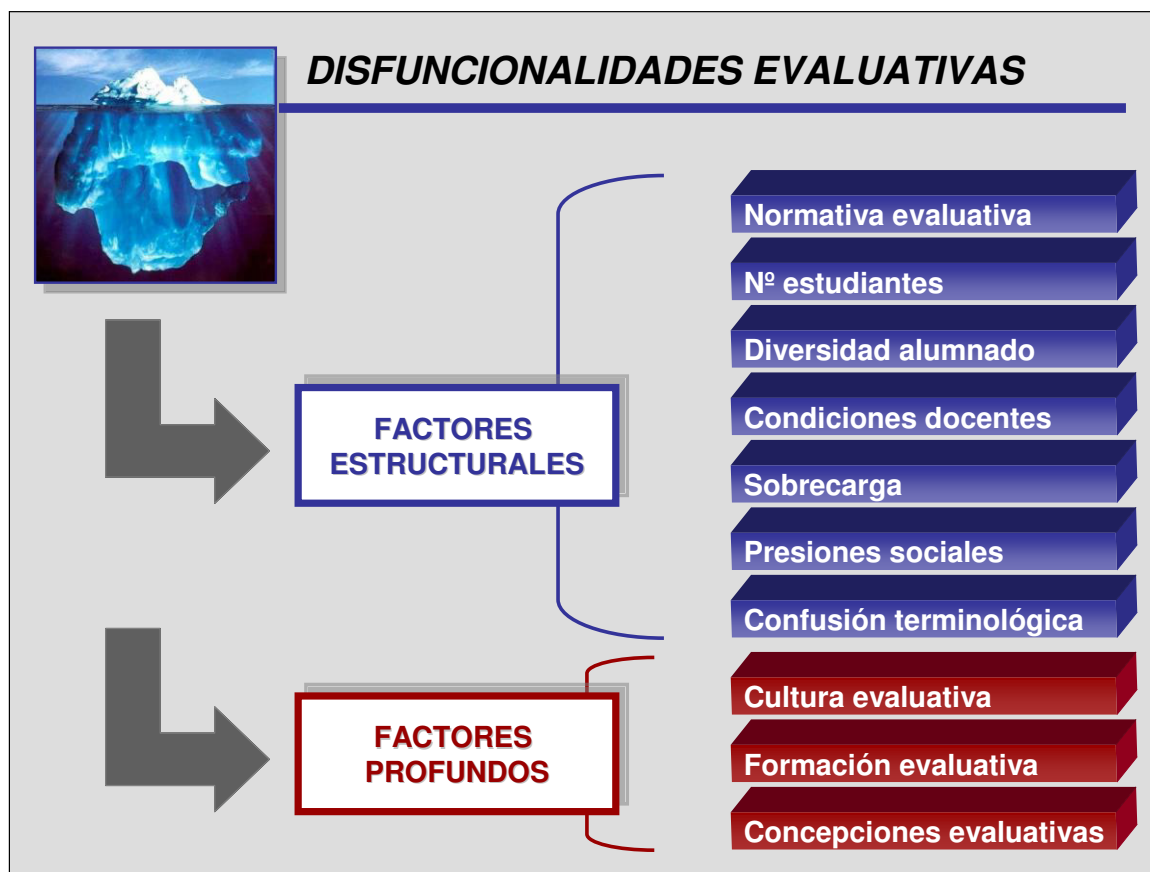


Figura 12. Causas profundas y estructurales de disfuncionalidad evaluativa.

Tanto si son profundas como estructurales, las causas representadas en la figura 12 guardan entre sí una estrecha relación de interdependencia. Su poder de influencia en las deficiencias detectadas es desigual, pero todos ellos deben ser valorados para lograr un avance decidido en las prácticas evaluativas universitarias. En opinión de Sanmartí y Jorba (1995) o Ahumada (2002), la desconsideración de algunas de estas circunstancias es lo que conduce al fracaso de la mayor parte de los procesos de renovación en el ámbito de la evaluación.

3.1 Condiciones o factores estructurales que originan disfuncionalidades evaluativas y/o favorecen su persistencia.

Afirma A. Brew (2003) que “uno de los grandes desafíos para todo el profesorado es introducir evaluaciones realistas y relevantes dentro de las limitaciones de los recursos disponibles o enfrentándose a actitudes tradicionales” (p. 188). De acuerdo con Ahumada (2002),

continuar asegurando que la evaluación es un proceso continuo, dinámico, flexible, participativo, reflexivo e inherente a todo aprendizaje, podría llegar a constituirse en un discurso estéril ya que las condiciones en que se realiza el proceso docente actual no permiten cumplir con estas utópicas propiedades del proceso evaluativo, que están, sin embargo, expresadas reiteradamente en nuestros escritos y comunicaciones. (p. 34).

Desarrollados a continuación, los factores estructurales anticipados en la figura 12 representan circunstancias que dificultan la realización de evaluaciones de calidad y rigurosas, pero no lo impiden directamente (Santos, 1999).

3.1.1 La influencia de las normas administrativas y reglamentos.

Las normas administrativas y reglamentos a menudo contienen regulaciones que van en contra de los principios pedagógicos en los que debería inspirarse la evaluación (Ahumada, 2005; Brown, 2003b; Glasner, 2003). Las condiciones detalladas a continuación figuran entre las más reportadas a la hora de señalar factores que dificultan la praxis evaluativa en la universidad:

- *El calendario prefijado de evaluación y los plazos de exámenes.* Por lo general, las normas y reglamentos que regulan el ejercicio de la evaluación en las universidades determinan un calendario prefijado de periodos y momentos para su ejercicio (Ibarra & Rodríguez, 2010; Padilla et al., 2010; M. L. Pérez et al., 2000; Villa & Poblete, 2011). Esta rigidez, no permite dar respuesta a la diversidad de ritmos de aprendizaje que se produce entre el alumnado (Bain, 2007; Brown, 2003a; Rizo, 2004). Según Sanmartí (2007), forzar a las/os estudiantes a poner a prueba su aprendizaje en un momento externamente determinado incrementa la probabilidad de fracaso. “Calificar demasiado pronto en el proceso de aprendizaje puede conducir a los estudiantes a la dañina inferencia de que ser bueno significa no tener que intentarlo, y que tener que intentarlo significa no ser hábil en la materia” (J. Chappuis, 2014, p. 26). Atendiendo a este mismo razonamiento, la QAA (2014) orienta a las universidades del Reino Unido a favorecer una aproximación flexible al tiempo de la evaluación para no perjudicar el aprendizaje discente.

- *La modularización o fragmentación curricular.* La excesiva compartimentalización del conocimiento desde la que se elaboran los planes de estudio, condicionan la forma de llevar a cabo la evaluación (Porto, 2006). Así lo advierten también Knight (2005) y Gibbs y Simpson (2009), que ven en la modularización una de las causas que explica el mantenimiento de prácticas evaluativas descontextualizadas y simplificadoras del aprendizaje.
- *Descoordinación en materia evaluativa.* Unido a la fragmentación curricular, la evaluación se lleva a cabo de forma individual, presumiendo que incumbe sólo a la/al docente de cada módulo/asignatura (Briggs et al., 2003; Gairín, 2009; Sayós, 2010). A juicio de Porto (2008, 2010), esta descoordinación representa una limitación que dificulta el desarrollo de prácticas evaluativas de calidad.
- *Otras regulaciones.* Además de los factores recopilados en las líneas anteriores, los reglamentos y normativas refuerzan una visión de la evaluación que favorece la persistencia de muchas de sus disfuncionalidades. Por una parte, se aprecia una reducción evidente de sus funciones, confirmando una de las limitaciones evaluativas más reiteradas:

En ningún momento se hace mención a la evaluación como un mecanismo para la mejora del aprendizaje. Más bien al contrario, el análisis de la normativa nos lleva a pensar que no siempre se diferencia entre evaluación y calificación, y que esta última es esencial para discernir si un alumno ha superado o no un período formativo, cuestión importante a cuya resolución se orienta todo el sistema de evaluación. (Padilla et al., 2010, p. 11).

Por la otra, “los estatutos universitarios presentan un mayor interés por los instrumentos utilizados que por los criterios o tipos de evaluación” (Ibarra & Rodríguez, 2010, p. 394), entre los que tienen una especial presencia las pruebas o exámenes, globales o parciales.

3.1.2 El incremento del número de estudiantes o la masificación de las aulas universitarias.

El elevado número de estudiantes con el que habitualmente se trabaja en la universidad representa uno de los factores más señalado a la hora de explicar las

disfuncionalidades evaluativas y su persistencia (AR@HK, 2009d; Cano & Ion, 2012; Capllonch et al., 2009; Gutiérrez et al., 2013; Morales-Olivas & Estañ, 2007; o Zaragoza et al., 2009). El empleo de procedimientos evaluativos cuantitativos, el predominio de la evaluación final y sumativa o la ausencia de feedback o valoraciones formativas durante el proceso son algunas de las repercusiones plausibles de este factor (Glasner, 2003; Rizo, 2004; M. P. Romero, 2004).

3.1.3 La creciente diversidad del alumnado universitario.

El alumnado universitario responde a un perfil cada vez más amplio y diverso, que hace coincidir en un mismo grupo circunstancias y trayectorias personales, académicas y profesionales muy distintas (Moore & Murphy, 2009; Zabalza, 2002). Este incremento del grado de diversidad representa otra dificultad a la hora de mejorar los sistemas de evaluación en opinión de Letelier (2009) o Gibbs y Simpson (2009).

En este marco de heterogeneidad convergen circunstancias y necesidades particulares muy diversas. A efectos evaluativos resulta especialmente visible el caso del alumnado matriculado a tiempo parcial por distintas circunstancias (Capllonch et al., 2009; V. M. López et al., 2007). La generalización del principio de formación a lo largo de la vida hace prever una mayor concurrencia de alumnado adulto en las aulas universitarias, dando lugar a nuevas necesidades evaluativas (itinerarios paralelos o alternativos, acreditación de aprendizajes y experiencias profesionales, etc.).

3.1.4 La incidencia de otras dificultades asociadas a la docencia en la evaluación del aprendizaje.

Desarrolladas en el capítulo primero, las dificultades generales asociadas a la docencia universitaria representan otro factor explicativo de las deficiencias observadas en la evaluación y su resistencia a la mejora (Porto, 2010). Estas condiciones conforman el contexto real para el ejercicio de la docencia e inciden en el desarrollo de todas las funciones vinculadas a ella, incluida la evaluación. Aspectos como la presión investigadora, la sobrecarga, la precariedad laboral o la falta de recursos representan factores que pueden condicionar la calidad de la praxis evaluativa.

3.1.5 La sobrecarga docente y discente.

La superación de los rasgos de disfuncionalidad descritos exige avanzar hacia fórmulas evaluativas pedagógicas y continuas, sensibles con la construcción del aprendizaje, y así será planteado en un próximo apartado. Sin embargo, en comparación con las prácticas puntuales y finalistas, la evaluación pedagógica y continua tiende a sobrecargar a estudiantes y docentes (Capllonch & Buscà, 2012; Coll et al., 2007; A. López et al., 2007; Rué, 2007).

Existen testimonios que defienden que este incremento de la carga de trabajo es asumible desde las dedicaciones reconocidas a docentes y estudiantes (R. Romero et al., 2014), pero no se trata de una postura generalizada. Sanmartí (2007), Sabaté et al. (2008) o Forés et al. (2014) señalan que la superación de sistemas evaluativos finalistas y puntuales, por lo general, supone para el alumnado la acumulación de mayor cantidad de tareas y la reducción de los plazos para su realización. Para el profesorado, incrementa sustancialmente el volumen de correcciones (Morales-Oliva & Estañ, 2007; Navarro et al., 2010; M. P. Romero, 2004; Zaragoza et al., 2009) y lleva a invertir mayor tiempo en las tareas de registro, feedback y seguimiento asociadas.

3.1.6 Presiones sociales.

“Nos guste o no, el producto más significativo de la educación superior son las calificaciones que obtienen los estudiantes, más que la calidad de su experiencia de aprendizaje” (Race, 2003, p. 78). En opinión de Santos (2001, 2003) esta presión social por el resultado ejerce un importante poder de influencia en las prácticas evaluativas que tienen lugar en la universidad. Por una parte, explican la búsqueda de fórmulas de calificación que se ajusten claramente a la lógica cuantitativa, a pesar de que simplifiquen y descompongan en partes el aprendizaje. Por la otra, la necesidad de responder simultáneamente a las expectativas de las diversas audiencias interesadas (agentes empleadores, familias, asociaciones profesionales, las propias instituciones formativas, etc.) puede representar un obstáculo para introducir innovaciones en materia evaluativa según Crossouard (2010) y McDowell y Sambell (2003). Estas audiencias emplean la evaluación como mecanismo de selección, según el recuerdo y las concepciones inscritas en su propia experiencia; en consecuencia, tienden a ser reacias a la introducción de innovaciones (Race, 2003).

3.1.7 Confusión o ambigüedad terminológica.

Tratada en el capítulo anterior, la polisemia o confusión terminológica a menudo representa un obstáculo a la hora de emprender propuestas o proyectos de mejora en el terreno de la evaluación (JCSEE, 2007; G. Rodríguez & Ibarra, 2012). La ambigüedad da lugar a un contexto de confusión e incompreensión mutua, que genera inseguridad y dificulta procesos compartidos de puesta en común o intervención. “La paradoja es que el profesor que debe resolver la evaluación contempla abrumado este desconcierto terminológico como espectador de su propio espectáculo sobre un escenario en el que actúa” (J. M. Álvarez, 2003, p. 316).

El origen de tal confusión se explica por la confluencia de dos grandes circunstancias: la falta de formación o capacitación evaluativa específica del profesorado y la prolífica evolución descrita para el propio término. Asimismo, atribuimos su mantenimiento a la unión de varias circunstancias que ya han sido reseñadas; entre las más evidentes, destaca la defensa de la evaluación como espacio de actividad privada y la consiguiente descoordinación en las prácticas evaluativas.

3.2 Condiciones o factores profundos que originan disfuncionalidades evaluativas y/o favorecen su persistencia.

Calificadas como profundas, las variables incluidas en este segundo grupo no se limitan a condiciones estructurales o contextuales y lindan a menudo con constructos complejos como las creencias o representaciones profundas, la tradición o la cultura.

La revisión de las aportaciones recopiladas en la literatura específica sobre el tema, nos ha llevado a identificar tres factores profundos o no estructurales que dificultan especialmente la evaluación del aprendizaje universitario y su mejora (figura 12).

3.2.1 La falta de cultura evaluativa en el sistema educativo en general y en el universitario, en particular.

En el sistema educativo español existe una falta transversal de cultura evaluativa que afecta a estudiantes y profesorado (Letelier, 2009). Entendiendo por cultura “el conjunto de significados compartidos por un conjunto de personas a través de un tiempo de

experiencia y de intercambios psicosociales” (Santos, 2003, p. 39), dicha afirmación describe una falta general de costumbre e incidencia real de la evaluación en los distintos niveles educativos.

Los indicios reportados en la literatura refuerzan la tesis de una cultura evaluativa pobre y escasamente desarrollada. Algunos de los rasgos que así lo evidencian son:

- *Escasa cultura general de la revisión.* Zabalza (2003) explica este rasgo como una predisposición a la repetición, a la improvisación o a iniciar cada proceso formativo como si fuese independiente del anterior. Afecta a la evaluación infrecuente de los distintos componentes del proceso formativo y explica también la escasa incidencia de la metaevaluación.
- *Escasa cultura general de documentación.* Ligado al primer rasgo, rara vez se recogen y registran datos y evidencias fidedignas sobre las que apoyar la revisión y mejora durante el proceso (Zabalza, 2003).
- *Cultura evaluativa inmovilista.* Por todo lo anterior, las prácticas evaluativas se tiñen de un cierto inmovilismo y conservadurismo (Clipa & Duta, 2010; Serrano, 2010).

No es posible que mientras la dinámica mundial, impulsada por la globalización y la tecnología, genere cambios permanentes en los distintos ámbitos del quehacer humano, la práctica evaluativa y su formación en este campo no esté en consonancia con esta dinámica ni responda a las nuevas tendencias educativas. (Smither, 2006, p. 156).

- *Cultura evaluativa escasamente participativa.* La intervención que se genera en torno al fenómeno evaluativo tiende a ser repetitiva y limitada. Por lo general, la actividad evaluadora recae única y exclusivamente en el profesorado, de modo que se genera una escasa cultura participativa entre las/os estudiantes (I. Álvarez, 2008a; Castillo, 2004; Gil-Flores, 2012).
- *Cultura evaluativa individualista.* Ligado al rasgo anterior, la cultura evaluativa en la universidad da muestras de individualismo en la intervención docente y discente. El profesorado diseña e implementa los dispositivos evaluativos en solitario y sin coordinación y el aprendizaje del alumnado se valora por medio de sistemas meritocráticos que atienden principalmente a logros individuales.
- *Cultura evaluativa opaca o secretista.* Perrenoud (2010) señala que la autonomía o libertad de cátedra del profesorado universitario, llevada a extremo, genera un cierto

secretismo en torno a la evaluación. Junto al reseñado individualismo, este secretismo dificulta el acceso a las diversas prácticas evaluativas para su revisión, coordinación o mejora.

- *Cultura evaluativa conceptual y declarativa.* Este rasgo hace referencia a la atención preferente de la evaluación a saberes declarativos o contenidos conceptuales, en detrimento de otros componentes del aprendizaje.

Las manifestaciones destacadas son producto de la expansión de prácticas evaluativas disfuncionales y, al mismo tiempo, obstaculizan el avance y la mejora, por eso se incluyen en este apartado.

3.2.2 El problema de la formación en evaluación.

Stiggins (1993/2012) afirma que la capacitación evaluativa representa una negligencia a superar en la formación docente. Esta deficiencia es consecuencia de la indefinición profesional descrita para la docencia universitaria en el primer capítulo, y conecta con un debate más amplio sobre su insuficiente preparación didáctico-pedagógica (Margalef & Álvarez, 2005; Montero, 2006). Su incidencia se encuentra ampliamente suscrita tanto a nivel nacional, como internacional (B. Álvarez et al., 2007; Banta, 2007; Erwin, 2003; Nieto, 2000; Porto, 2005b; Stiggins, 1999; Trillo & Porto, 1999; Villardón, 2006). En la práctica, se materializa en propuestas formativas insuficientes e inadecuadas, que no garantizan la capacitación requerida para el desempeño de las funciones evaluativas, ni la superación de las disfuncionalidades advertidas en la misma.

En cuanto a la insuficiencia, Mas (2011, 2012) y Santos (2003) cuantifican la disminución que se produce en la formación básica docente en la medida que aumenta el nivel educativo en el que vaya a ejercerse. Ideada para propiciar la preparación didáctico-pedagógica necesaria (incluida la capacitación evaluativa), ésta es de 240 ECTS para las/os maestras/os de Educación Infantil y Primaria y de entre 60 y 120 ECTS para el profesorado de Educación Secundaria²³. Estos programas formativos no dedican atención suficiente al desarrollo de la competencia evaluativa en opinión de muchas/os expertas/os (Mertler, 2009; Sanders & Vogel, 1993/2012; Stiggins & Conklin, 1988; Wise & Lukin, 1993/2012), pero en la enseñanza universitaria la situación empeora. En ese caso, el acceso a la

²³ En medio de este debate, Montero (2006) y J. M. Esteve (2006) dejan planteadas las repercusiones de los modelos de formación docente simultáneos (Educación Infantil y Primaria) y consecutivos (enseñanza secundaria) en la enseñanza y la identidad profesional del profesorado.

docencia no está sujeto a ninguna capacitación pedagógica específica, a menos que la propia institución así lo exija por medio de programas propios de formación inicial. La formación permanente tampoco parece compensar esa laguna en la capacitación inicial: la evaluación del aprendizaje no figura entre los principales centros de interés y son poco frecuentes las actividades de formación continua que profundizan en ella (V. Álvarez et al., 2011; Elexpuru, Martínez, Villardón, & Yániz, 2006; JCSEE, 2007; R. McDonald et al., 2000).

El déficit de formación evaluativa inicial y permanente predispone al profesorado a reproducir los patrones clásicos de evaluación (B. Álvarez et al., 2007; Hortigüela et al., 2015; V. M. López & Palacios, 2013; Monereo, 2013). Las demandas evaluativas cotidianas se resuelven recurriendo a modelos conocidos, lo que genera un doble efecto: además de contribuir a la perpetuación de las disfuncionalidades advertidas, la reproducción disminuye en el profesorado la sensación de incapacidad y, con ello, la necesidad formativa (Tejedor, 1999).

En cuanto a la inadecuación, la oferta formativa para la evaluación del aprendizaje parece ensombrecida por una falta de conexión con la realidad del aula y las necesidades del profesorado (Gullickson, 1993/2012; Stiggins & Conklin, 1998; Zeichner, 2010). Su desarrollo se ha centrado prioritariamente en incrementar el repertorio de técnicas y herramientas, pero sin llegar a modificar las representaciones profundas sobre la evaluación y los principios básicos sobre los que se asienta (Knight, 2005; Monereo & Castelló, 2009b; Verdejo, 2008). Sin embargo, en cualquier área o competencia que desee desarrollarse, la formación no resulta efectiva a menos que afecte a la identidad docente (Monereo, Weise, & Álvarez, 2013). Dicha identidad está formada por ideas subjetivas sobre su quehacer, sus funciones y sus competencias y condiciona el modo de entender la calidad profesional (Martínez de la Hidalga & Villardón, 2014).

Estas concepciones o creencias profundas sobre la naturaleza de la evaluación y su resistencia al cambio, representan la tercera gran dificultad asociada a la mejora de las prácticas evaluativas en la universidad y sobre ella nos detendremos a continuación.

3.2.3 Las creencias, teorías, concepciones o representaciones profundas sobre la evaluación y su resistencia al cambio.

Las disfuncionalidades que actualmente se observan en los sistemas evaluativos universitarios se relacionan con las concepciones educativas profundas que se sostienen sobre el aprendizaje, la enseñanza y la evaluación (Brown, 2003b; Knight, 2005; Mateos &

Pérez, 2006). Dichas concepciones condicionan la interpretación que realizan estudiantes y profesorado de las distintas situaciones formativas y, en consecuencia, su forma de actuar (De la Cruz et al., 2006; Korthagen, 2010; D. Salinas, 2002).

Las concepciones son un constructo que los investigadores han creado para referirse a parte del conocimiento personal que los seres humanos poseen. En pie de igualdad con este término, ha aparecido en la literatura relacionada con el pensamiento de los padres, profesores, etc., un conjunto de expresiones que principalmente tienen en común con él su marcado carácter de conocimiento personal y a la vez social, así como su gran importancia de cara a los procesos de influencia educativa. (Buendía et al., 1999, p. 130).

Creencias, representaciones o teorías implícitas son algunos de los términos empleados como sinónimos en referencia a este constructo (Pajares, 1992). Aunque existan matices y leves diferencias entre todos ellos, a los propósitos de la presente investigación interesan los rasgos que mantienen en común: su componente cognitivo y su influencia en la manera de interpretar las situaciones educativas y evaluativas (Buendía et al., 1999; Pajares, 1992; Ponte, 1994; M. Prieto & Contreras, 2008).

Las concepciones educativas profundas²⁴ pueden formarse sobre cualquiera de los elementos que interceden en el proceso formativo. En este trabajo resultan de especial interés aquellas relativas a la evaluación, pero debe considerarse la estrecha vinculación de las mismas con las representaciones generales conformadas sobre el aprendizaje y la enseñanza (De la Cruz et al., 2006; R. López, 2001; Smitter & Siso, 2007).

Las concepciones evaluativas profundas son aprendizajes inconscientes, contruidos por medio de experiencias de evaluación acumuladas de forma no deliberada (Buendía et al., 1999; De la Cruz et al., 2006; Pajares, 1992). Representan un concepto de importancia capital porque condicionan la forma de posicionarse ante la evaluación de estudiantes y docentes. Cada una de las decisiones evaluativas del profesorado parte de estas creencias, y así también las que adopta el alumnado como principal agente evaluado.

²⁴ En la presente investigación el calificativo profundo se añade, en referencia a su origen y naturaleza (surgen de experiencias personales y vicarias y permaneces implícitas). En el extremo opuesto, las representaciones, creencias o concepciones explícitas hacen referencia a aprendizajes conscientes, derivados de la exposición y contacto con modelos y principios teóricos, más que a vivencias directas.

Desde esta perspectiva, las creencias o teorías profundas sobre la evaluación estarían en la base de muchas de las deficiencias o disfuncionalidades señaladas. En la misma medida, la mejora de la evaluación del aprendizaje quedaría supeditada al desarrollo de nuevas concepciones que den lugar a nuevas prácticas (Crossouard, 2010; Knight, 2005; R. McDonald et al., 2000). Sin embargo, el origen y la naturaleza de estas últimas las hacen extremadamente resistentes al cambio, dificultando el avance y la mejora.

Para comprender el papel que juegan las representaciones profundas en la existencia y persistencia de las disfuncionalidades evaluativas, los próximos apartados recorren sus principales rasgos característicos y su poderoso arraigo. La tabla 21 ofrece un anticipo de las características más relevantes para comprender el papel que juega este constructo, según lo precisado en los próximos subapartados.

Tabla 21

Resumen de los rasgos básicos de las concepciones evaluativas profundas

CONCEPCIONES EVALUATIVAS PROFUNDAS	
SON...	Creencias, teorías o representaciones implícitas sobre evaluación, estrechamente vinculadas a aquellas otras referidas al aprendizaje y a la enseñanza.
SU FUNCIÓN...	Representan el marco de referencia sobre el cual docentes y discentes adoptan decisiones e intervienen en las distintas situaciones formativas, especialmente aquellas que hacen referencia expresa a la evaluación.
SUS RASGOS...	• Origen implícito. • Naturaleza implícita. • Independencia de las representaciones o teorías explícitas. • Consecuencia de aprendizajes informales. • Carácter encarnado, personal y emocional. • Naturaleza automática. • Carácter contextual. • Fuertemente resistentes al cambio: ✓ Difíciles de identificar y explicitar. ✓ Arraigo (muy personales, encarnadas). ✓ Exigen todo un proceso de reconstrucción. ✓ Independencia de las teorías o representaciones explícitas.

Nota: Elaboración propia a partir de las contribuciones de Buendía et al. (1999), R. López (2001), Marín y Cervi (2006), Mateos y Pérez (2006), Pajares (1992), Ponte (1994), Pozo et al. (2006) y Reber (1993).

3.2.3.1 Origen y características básicas de las representaciones evaluativas profundas.

La revisión de las aportaciones que diversas/os autoras/es realizan en torno a las creencias o concepciones profundas, nos lleva a identificar siete rasgos básicos que condicionan su influencia y funcionamiento (ver tabla 22).

Tabla 22

Descripción de los rasgos básicos de las concepciones evaluativas profundas

RASGOS BÁSICOS DE LAS CONCEPCIONES EVALUATIVAS PROFUNDAS	
DE ORIGEN IMPLÍCITO	Tienen lugar por medio de un proceso de construcción implícito, a partir de la exposición reiterada a ciertos patrones. Este proceso de construcción es involuntario: tiene lugar sin que exista una intención o deseo expreso por aprender y sin consciencia clara sobre aquello que se aprende.
DE NATURALEZA IMPLÍCITA	Son de naturaleza implícita e inconsciente y, por tanto, difícilmente comunicables y accesibles a la conciencia. Su análisis e investigación resulta muy compleja porque el solo intento de verbalizarlas supone un paso en su transformación. El contraste de perspectivas o perspectivismo (Rodrigo, 1993), la participación activa y la reflexión, son algunos de los cauces que favorecen su explicitación.
INDEPENDIENTES DEL SABER EXPLÍCITO	Son independientes de la teoría explícita sobre la evaluación. En consecuencia, pueden producirse discrepancias entre lo que se sabe o defiende de forma racional y lo que se muestra en la práctica.
FRUTO DE APRENDIZAJES INFORMALES	Frente a los aprendizajes explícitos, que se enseñan y desarrollan en contextos formativos formales, son producto de aprendizajes informales. Se aprenden sin intención expresa y no se enseñan; aunque puedan desarrollarse en el ámbito educativo formal, no responden al currículum explícito ni son producto de instrucción programada.
ENCARNADAS	Frente al saber declarativo, teórico y explícito, representan el saber de la experiencia y la vivencia. En consecuencia, tienen un fuerte componente emocional y carácter encarnado.
DE NATURALEZA AUTOMÁTICA	Frente al carácter deliberado de los aprendizajes explícitos, se activan de forma automática o no controlada. Representan herramientas funcionales, al permitir responder de forma rápida a la complejidad de las situaciones que plantea la práctica sin apenas consumir recursos.
DE CARÁCTER CONTEXTUAL	Son aprendizajes contextuales, estrechamente ligados a las condiciones en las que se originaron. En situaciones conocidas desempeñan una función pragmática para la actuación y la toma de decisión. En contextos o situaciones novedosas, no se produce la activación automática a la que hacía referencia el rasgo anterior.
RESISTENTES AL CAMBIO	Aunque pueden modificarse, sus rasgos propios las hacen menos flexibles y dinámicas que otros aspectos del conocimiento.

Nota: Síntesis elaborada a partir de las contribuciones de fuentes diversas (Buendía et al., 1999; J. Contreras, 2010; A. I. Gómez, 2010a; Korthagen, 2010; R. López, 2001; Marín & Cervi, 2006; Mateos & Pérez, 2006; Pajares, 1992; Ponte, 1994; Pozo et al., 2006; Reber, 1993; Rodrigo, 1993, 1997).

3.2.3.2 *El arraigo de las representaciones evaluativas profundas o su resistencia al cambio.*

Uno de los rasgos que caracteriza las representaciones profundas configuradas sobre la evaluación es su resistencia al cambio.

Frecuentemente cuando tratamos de innovar en el aula nos tropezamos con las concepciones que los alumnos y las alumnas han ido desarrollando a lo largo de su trayectoria académica. Y no estamos hablando de sus ideas alternativas, sino de esas rutinas que han asimilado respecto a lo que es aprender y evaluar. (Otero, 2000, p. 151).

Aunque las palabras anteriores hagan referencia al colectivo discente, las representaciones profundas del profesorado también representan un obstáculo para la mejora y el cambio por su escasa alterabilidad (Ahumada, 2005; Monereo & Castelló, 2009b). La tabla 23 detalla las principales causas que explican tal arraigo, partiendo de las aportaciones de varias fuentes.

Tabla 23

Resistencias al cambio de las concepciones evaluativas profundas: descripción de causas

Las concepciones evaluativas profundas son altamente resistentes al cambio porque...	
EL CAMBIO EXIGE EXPLICITACIÓN	Aunque tengan naturaleza implícita, sólo pueden reconstruirse previa explicitación. Su modificación exige hacerlas conscientes y ello entraña serias dificultades. Previamente señaladas, el intercambio de perspectivas, la participación y la reflexión constituyen las más poderosas herramientas para favorecer la explicitación. Sin embargo, las acciones formativas diseñadas para el cambio evaluativo no suelen propiciarlas; tradicionalmente adoptan una orientación individual, no provocan la confrontación de puntos de vista y se centran en sesiones expositivas, en las que tampoco se potencia la participación²⁵, ni se establecen procesos de reflexión a partir de la experiencia.

²⁵ No, al menos, en la medida en la que ésta podría producirse por medio de otros procedimientos como la resolución de problemas o el análisis de casos, entre otras opciones.

Las concepciones evaluativas profundas son altamente resistentes al cambio porque...

SU CARÁCTER ENCARNADO DIFICULTA EL CAMBIO	Su vinculación con la experiencia les dota de una alta significatividad: • Las experiencias y vivencias acumuladas tienen mayor peso e influencia en las creencias profundas que las palabras, medio en el que habitualmente se apoyan las iniciativas formativas destinadas al cambio de concepciones. • El componente emocional y personal de las concepciones evaluativas profundas, provoca que no puedan abandonarse nunca del todo; como se expone a continuación, eso imposibilita que el cambio se resuelva por mera sustitución.
EXIGEN UN PROCESO DE RECONSTRUCCIÓN	Su cambio exige un complejo proceso de reconstrucción y reorganización, de modo que las representaciones vinculadas a la experiencia informal, se integren en aquellas otras asociadas a los modelos teóricos y fundamentos epistemológicos de referencia. La modificación no se produce por yuxtaposición de nuevas creencias, sino que tiene que tener lugar el cuestionamiento de las previas (reto complejo por su carácter encarnado).
SU INDEPENDENCIA DEL SABER EXPLÍCITO	Las representaciones explícitas e implícitas responden a funciones diferentes, por eso pueden coexistir, incluso siendo contradictorias. Mientras que las representaciones o aprendizajes explícitos responden a una función epistémica (obtener respuestas, comprender y dar sentido), las creencias implícitas tienen una función pragmática (evitar problemas y tener éxito). Desde esta perspectiva, la exposición a nuevos referentes sobre evaluación no asegura el abandono de los construidos por medio de la experiencia implícita, ni mucho menos su puesta en práctica. Sin embargo, las acciones formativas diseñadas para el cambio de concepciones en docentes y discentes se construyen presuponiendo que la relación entre teoría y práctica es lineal, y confiando en el poder de la explicación y la palabra.

Nota: Síntesis elabora a partir de las contribuciones de fuentes diversas (Korthagen, 2010; V. M. López, 2004; Marín & Cervi, 2006; Mateos & Pérez, 2006; Pozo et al., 2006; M. Prieto & Contreras, 2008).

4. Propuestas y sugerencias de mejora para la evaluación del aprendizaje universitario

A lo largo de los tres últimos apartados han sido expuestos los principales rasgos de disfuncionalidad que presenta el sistema tradicional de evaluación del aprendizaje universitario, los efectos que ocasionan en alumnado y profesorado, así como los factores que explican tales deficiencias y su persistencia. Desde este marco de referencia, el nuevo apartado ofrece una síntesis de las principales sugerencias que se proponen para la mejora de praxis evaluativa en dicho contexto formativo.

Desde el punto de partida dibujado, entendemos que mejorar la evaluación del aprendizaje significa introducir cambios y ajustes en aras a responder con eficacia a las funciones que le han sido atribuidas en el capítulo segundo. Esta afirmación incluye la

función social o acreditativa, pero afecta especialmente a las funciones pedagógicas, por ser las menos atendidas tradicionalmente y ofrecer la base para el ejercicio de la evaluación para el aprendizaje. Igual que Black y Wiliam (1998a, 1998b, 2009), Gibbs & Simpson (2009) también priorizan la mejora de la dimensión pedagógica de la evaluación por delante de su misión acreditativa.

No estamos defendiendo una evaluación poco fiable, pero sí que las formas de evaluación deben diseñarse primordialmente para fomentar un aprendizaje fructífero a lo largo de la vida [...]. Los niveles subirán gracias a la mejora del aprendizaje de los estudiantes y no tanto por la mejora de la medición de un aprendizaje limitado. (p. 10).

Mejorar la evaluación exige desarrollar nuevas prácticas evaluativas y, ante todo, nuevas concepciones o formas de entenderla. Como afirma Crossouard (2010), “la pregunta de fondo es qué tipo de aprendizajes queremos para nuestras sociedades y a qué tipo de prácticas educativas debemos aspirar para apoyarlas” (p. 256). Probablemente por ello, la evaluación se presenta como la dimensión más resistente a las innovaciones en el contexto universitario (Margalef, 2009; M. L. Pérez, Juandó, & Carretero, 2009). Lograr que desempeñe con eficacia sus funciones y constituya una herramienta de indagación, que proporcione datos para la mejora del aprendizaje y su acreditación exige el desarrollo de un proyecto de revisión, ajuste e innovación, amplio y complejo. De acuerdo a la exposición de causas de disfuncionalidad efectuada, en el mismo deben contemplarse medidas para incidir sobre las variables estructurales y profundas que condicionan las prácticas actuales y favorecen su persistencia. Así lo contempla la figura 13, que anticipa las propuestas de mejora concretas para cada caso:

PROPUESTAS DE MEJORA

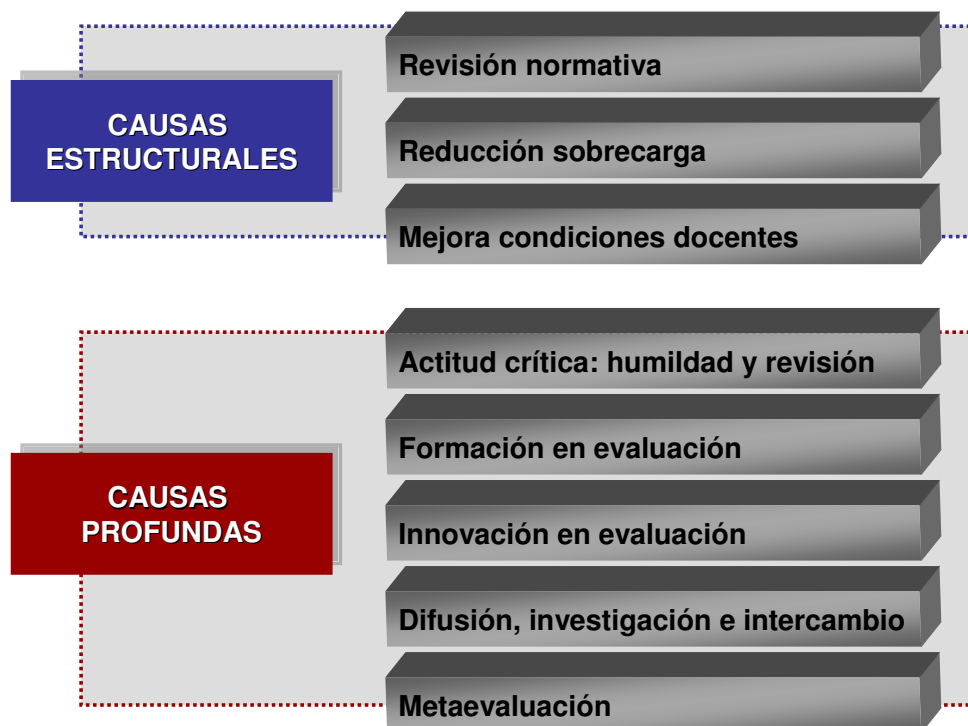


Figura 13. Propuestas para la mejora de la evaluación del aprendizaje en la universidad.

4.1 Iniciativas estructurales: sugerencias de mejora.

Las condiciones contextuales o estructurales configuran el contexto de la evaluación del aprendizaje e inciden en las prácticas que sobre el mismo se desarrollan. Su efecto repercute en los sistemas evaluativos empleados y mediatiza las intervenciones de mejora emprendidas. Sin embargo, por lo general hacen referencia a cuestiones que escapan al control docente y, en consecuencia, la investigación en curso no se detiene ni insiste excesivamente en ellas.

El análisis de las causas estructurales realizado en un apartado anterior, nos lleva a resumir tres iniciativas generales de mejora en este sentido:

4.1.1 Revisión de las normas y condiciones administrativas que condicionan la evaluación del aprendizaje en la universidad.

Esta primera línea de intervención reclama la revisión de las normas que regulan el ejercicio de la evaluación del aprendizaje para detectar y poner remedio a los aspectos que contribuyen al desarrollo y mantenimiento de prácticas disfuncionales (calendario predefinido, exigencias en cuanto a formatos, gestión de las convocatorias, etc.) (Ahumada, 2002; Cano & Ion, 2012; Ibarra & Rodríguez, 2010; Padilla et al., 2010). Smitter y Siso (2007) hablan de la dotación de soportes legales y técnicos adecuados, de modo que las imposiciones normativas regulen la función acreditativa de la evaluación sin interceder en la pedagógica.

La revisión de estos marcos reguladores también debe ofrecer respuesta a nuevas circunstancias y situaciones que se van generalizando en los últimos años y que repercuten en la evaluación. Ante la evidente diversificación del perfil discente universitario, por ejemplo, emerge la necesidad de incorporar a las normativas evaluativas el reconocimiento de aprendizajes previos efectuados en el ámbito académico, personal o profesional (Letelier, 2009; Quality Assurance Agency-QAA, 2014). La necesidad de reconocer diferentes itinerarios evaluativos, en atención a estudiantes con desigual dedicación a su formación, también representa una inquietud clásica, reactivada por el impulso del EEES a la evaluación continua (Capllonch et al., 2009; V. M. López et al., 2007; Navarro et al., 2010). Esta tendencia supondría reducir la estereotipia y uniformidad que acompaña a la evaluación en los marcos reguladores, dando lugar a nuevas alternativas que se ajusten mejor a las nuevas realidades²⁶.

4.1.2 Adopción de medidas para contrarrestar la sobrecarga que genera la masificación y la adopción de sistemas alternativos de evaluación.

Tomar distancia de los patrones evaluativos sumativos, finales y monometódicos y fomentar su funcionalidad pedagógica genera una sobrecarga que afecta tanto a discentes, como a docentes. Desde la perspectiva del profesorado, esta circunstancia se ha visto agravada en los últimos años con el incremento del número de estudiantes y la diversificación de los perfiles de acceso.

²⁶ La Ley 1/2013, de 10 de octubre, de Aprendizaje a lo Largo de la Vida del Gobierno Vasco, comparte el planteamiento de partida y procura un avance significativo en este sentido.

La literatura sugiere ciertas fórmulas o ajustes que afectan a la organización del trabajo para contrarrestar su efecto. Entre ellas:

- *Coordinación de tareas de evaluación y de ciclos de intensidad de trabajo.* La formulación de tareas evaluativas conjuntas entre las/os responsables de asignaturas o módulos diferentes puede constituir un recurso para reducir la sobrecarga docente y discente, además de favorecer la conexión interdisciplinar (Capllonch et al., 2009; B. Salinas & Cotillas, 2007). Para no perjudicar el aprendizaje discente con exigencias o presiones excesivas (Brown et al., 1996; Nicol, 2007, 2009), estas demandas deberían responder a un ciclo razonable de esfuerzo e intensidad de trabajo.
- *Reconocimiento de la carga docente.* La puesta en práctica de dispositivos evaluativos formativos y continuos exige al profesorado una dedicación mucho mayor que la requerida en la clásica evaluación final y sumativa. Para que tenga lugar su consolidación práctica, deberían introducirse medidas que favorezcan su reconocimiento laboral y profesional. A modo de ejemplo, Capllonch et al. (2009) proponen el desdoblamiento a partir de cierto número de matrículas o el reconocimiento del número de estudiantes como un factor de corrección en la estimación de la carga docente.
- *Tecnologías de la información y la comunicación.* El empleo de los recursos y las posibilidades que ofrecen las tecnologías de la información y la comunicación puede aliviar la sobrecarga que genera al profesorado la evaluación formativa y continua en grupos numerosos. Su uso resulta particularmente útil para favorecer el registro y seguimiento (Coll et al., 2007), así como en la generación y comisión de feedback (Evans, 2013; Lu & Law, 2012).
- *Implicación de las/os estudiantes en sistemas de coevaluación y autoevaluación.* Además de constituir una oportunidad formativa, la implicación del alumnado como agente evaluador puede contribuir a reducir la sobrecarga que genera al profesorado la tarea de corrección y comisión de feedback (Ballantyne, Hughes, & Mylonas, 2002; Carless, 2013a; Carless et al., 2011; Hounsell, 2007; Liu & Carless, 2006).

4.1.3 Mejora de las condiciones laborales y profesionales generales en la docencia universitaria.

El análisis de la enseñanza universitaria efectuado en el primer capítulo daba cuenta de una serie de condiciones laborales y profesionales que devalúan y dificultan el ejercicio de la docencia y, por extensión, de la evaluación.

Desde esta perspectiva, Banta, Lund, Black y Oblander (1996) afirman que uno de los principios que caracteriza la evaluación del aprendizaje de calidad es la consolidación de un ambiente de liderazgo y apoyo administrativo que procure las condiciones favorables para su desempeño profesional. De modo similar, Ion y Cano (2011), Stiggins y Duke (2008) y Stiggins (1999) sostienen que la mejora de la evaluación no depende exclusivamente del profesorado, y apelan a un liderazgo institucional que contribuya a su refuerzo y desarrollo.

4.2 Iniciativas no estructurales o profundas: sugerencias de mejora.

La resolución de las principales disfuncionalidades detectadas, así como el cambio de representaciones que debe tener lugar para la emergencia de nuevas prácticas evaluativas, no sólo responde a medidas de carácter estructural. Las cinco líneas de mejora expuestas a continuación están orientadas a reducir el impacto de los factores profundos en la praxis evaluativa. Estrechamente conectadas, este segundo bloque de sugerencias dibuja un marco de intervención mucho más amplio y complejo que el descrito para las variables estructurales:

4.2.1 Adopción de una actitud de humildad, revisión y autocrítica.

De acuerdo con Santos (2010b), la duda es una cualidad que caracteriza a las personas sabias y conduce a despojarse de toda certidumbre, incluida aquella que afecta a la práctica evaluativa. Así pues, mejorar los sistemas de evaluación del aprendizaje exige una actitud crítica y de autorrevisión, fundamentalmente por parte del profesorado, responsable inmediato de confeccionar y poner en práctica los dispositivos evaluativos (Bilbao, 2014; Race, 2003). “Los cambios, en la práctica pedagógica, no se generan sólo a

partir de decretos y leyes, sino afrontándolos con actitudes amplias, con disposición y conciencia, además de la preparación para asumir los retos planteados” (Smitter, 2006, p. 156).

El cambio y la mejora en la evaluación del aprendizaje pasa por el reconocimiento de la escasa autenticidad de las prácticas actuales y su revisión seria y rigurosa (Ahumada, 2005). Este ejercicio de revisión y reflexión hacen consciente al profesorado de la necesidad de cambio y potencia el hábito profesional de mejora continua (Villardón, 2006). Sin embargo, su desarrollo resulta especialmente delicado en el caso de la evaluación del aprendizaje, ya que en ella afloran creencias, temores, inercias y rutinas que suelen permanecer ocultas (Monereo & Castelló, 2009b).

La actitud de revisión y humildad que dinamiza los procesos significativos de reflexión es muy necesaria, pero no puede forzarse. Desde esta perspectiva, representa el prerequisite para emprender el resto de las iniciativas de mejora previstas. No es menos cierto que puede ser alimentada y estimulada en el desarrollo de las mismas, dando lugar a un proceso de retroalimentación.

4.2.2 Formación en evaluación para el profesorado universitario.

Fullan (2007) señala la formación como la mejor solución y, al tiempo, el mayor de los problemas en la cualificación docente. Así se constata también en el caso específico de la formación evaluativa, donde la inadecuada e insuficiente preparación del profesorado, no resulta fácil de resolver por medio de intervenciones formativas puntuales y compensatorias. Sweedler-Brown (1992) evidenció en su estudio que varias sesiones formativas no son suficientes para modificar hábitos fundamentados en años y que la consciencia sobre ciertas cuestiones no garantiza su puesta en práctica.

Para que la formación evaluativa sirva realmente como mecanismo para superar el marco de disfuncionalidades descrito, debe reunir una serie de requisitos asociados a su adecuación y suficiencia. Anticipados en la tabla 24, su descripción y justificación tiene lugar en los dos próximos apartados.

Tabla 24

Rasgos de la formación inicial y permanente del profesorado universitario para la mejora de los sistemas de evaluación del aprendizaje

La formación evaluativa del profesorado universitario debe...	
A	Ser suficiente , para lo que se precisa la combinación de iniciativas formativas iniciales y permanentes.
B	Orientarse al cambio en las representaciones profundas que el profesorado universitario sostiene sobre la evaluación, y no sólo a las cuestiones técnicas que inciden en dicho proceso.
C	Plantear escenarios y situaciones para promover deliberadamente el cambio en dichas representaciones...
	• Adoptando aproximaciones globalizadas e interdisciplinares que favorezcan la integración, en vez de la mera yuxtaposición. • Favoreciendo la participación activa del profesorado en formación para promover la explicitación de las creencias profundas sobre evaluación. • Potenciando el intercambio de perspectivas entre el profesorado en formación para promover la explicitación de sus representaciones evaluativas. • Estimulando en el profesorado en formación procesos de reflexión para favorecer la explicitación de las concepciones evaluativas implícitas. • Prestando especial atención a las disonancias que se producen entre creencias implícitas y explícitas del profesorado en materia evaluativa. • Planteando una formación práctica y contextualizada, basada en las situaciones y escenarios reales de la praxis evaluativa.
D	Promover el desarrollo de la competencia evaluativa del profesorado universitario...
	• Estimulando el desarrollo de conocimientos declarativos, procedimientos y actitudes para la movilización e integración armónica de los mismos. • Favoreciendo el desarrollo de la capacidad de transferencia a nuevas situaciones. • Preparando al profesorado para la complejidad y la incertidumbre que caracteriza a la práctica evaluativa.

4.2.2.1 Formación evaluativa cuantitativamente suficiente.

“Todas las evidencias disponibles apuntan a que en el futuro el profesorado deberá ser más docto en evaluación de lo que lo es en la actualidad, o de lo que lo ha sido en el pasado” (Stiggins, 1998, p. 1). En consecuencia, es indispensable contribuir a una preparación *suficiente* para el ejercicio de la competencia evaluativa, especialmente entre el profesorado universitario, tradicionalmente el menos formado desde la perspectiva didáctico-pedagógica. En correspondencia con las dificultades previamente referidas, el incremento de la oferta formativa en materia evaluativa representa una sugerencia de mejora muy extendida (Burgess, 2007; Ibarra & Rodríguez, 2010; JCSEE, 2007; Padilla et al., 2010; Villardón, 2006).

Para satisfacer esta necesidad es preciso desarrollar iniciativas de formación inicial y permanente centradas en la perspectiva pedagógica de la evaluación, y no tanto en la propia disciplina de referencia. Según Flemming (2003) la formación inicial destinada al profesorado principiante tendría mayores garantías de éxito que la permanente por no tener que enfrentarse al arraigo de ciertos hábitos y rutinas, pero ambas son relevantes y necesarias en las circunstancias actuales.

4.2.2.2 Formación evaluativa cualitativamente adecuada: la formación para el cambio de concepciones o representaciones profundas.

Las pocas iniciativas formativas que han abordado la evaluación del aprendizaje se han desarrollado desde una perspectiva técnica, muy centrada en las herramientas e instrumentos. Sin embargo, la superación de las disfuncionalidades identificadas exige llevar a cabo una intervención más amplia y profunda (Crossouard, 2010; Knight, 2005; Olmos & Rodríguez, 2010; Poblete, 2007; Sanders & Vogel, 1993/2012).

Construir un cierto bagaje técnico de nociones básicas ligadas a la evaluación, que garantice un repertorio bien nutrido de técnicas y procedimientos aplicables, representa una aportación indudablemente importante. Entre otros beneficios, este nivel de preparación podría contrarrestar la ambigüedad y confusión terminológica referida al tratar las causas de disfuncionalidad. Sin embargo, para que resulte verdaderamente efectiva y contribuya a la mejora, la formación debe llegar a modificar las concepciones que sostiene el profesorado sobre la evaluación (Bonsón & Benito, 2005; Cano & Ion, 2012; V. M. López, 2006, 2009b; Monereo & Castelló, 2009a, 2009b; M. L. Pérez et al., 2009).

Las creencias profundas representan el marco de decisión desde el que el profesorado plantea y afronta el proceso de evaluación por lo que, si no se modifican, no habrá un avance decisivo. Desde esta perspectiva, la formación inicial y permanente del profesorado puede entenderse como un proceso de cambio conceptual que impulse el avance desde las teorías evaluativas implícitas a otras científicamente más sustentadas (Martín, 2003). Así lo sostiene M. Martínez (2006) quien formula el reto de construir y consolidar esquemas de decisión en la práctica por medio de la formación inicial y permanente. El conocimiento explícito cambia con más facilidad que las representaciones implícitas, de modo que “puede crear, por así decirlo, nuevas zonas de desarrollo próximo para nuestras prácticas de aprendizaje y enseñanza” (Pozo et al., 2006, p. 104).

La modificación de las creencias evaluativas profundas representa una empresa ardua y compleja. Exige la creación deliberada de escenarios y situaciones que favorezcan

el cambio de representaciones, partiendo de la consideración explícita y pormenorizada de sus rasgos propios (Mateos & Pérez, 2006; Pozo et al., 2006).

Partiendo de la revisión realizada en el tercer apartado (tabla 21, tabla 22 y tabla 23), presentamos a continuación una serie de pautas u orientaciones generales para incrementar la probabilidad de que la formación docente inicial o permanente llegue a alterar las concepciones profundas sobre la evaluación del aprendizaje.

- La formación debe diseñarse desde un *enfoque interdisciplinar o globalizado*. Frente a las acciones formativas puntuales, consecutivas y fragmentadas, que separan la preparación disciplinar de la psicopedagógica (J. M. Esteve, 2006; I. Fernández & Márquez, 2014; Montero, 2006), parecen más propicios los planteamientos globalizados e interdisciplinares, que favorecen la integración que requiere el cambio de creencias.
- La formación debe provocar la *explicitación de las concepciones evaluativas profundas* para que pueda tener lugar su modificación. Algunas de las medidas que puede poner en juego para ello son:
 - ✓ Promover la *participación activa de los docentes cuyas concepciones se pretenden modificar*. Frente a los clásicos modelos formativos magistrales, que sitúan al agente destinatario en un rol fundamentalmente pasivo, conviene recurrir a metodologías más activas y participativas como el análisis de casos o la resolución de problemas (Mateos & Pérez, 2006). En opinión de Mateo (2012), esta participación debe tener lugar desde la fase de planificación de la actividad formativa, de modo que se optimice el aprovechamiento de la misma y sea posible la satisfacción de las distintas necesidades.
 - ✓ Promover el *intercambio o contraste de perspectivas*. Las iniciativas formativas deberían distanciarse del clásico enfoque individualista y potenciar experiencias de trabajo compartido e intercambio (Blanch & Corcelles, 2014; Mauri, Clarà, Ginesta, & Colomina, 2013; Prebble et al., 2004; Southwell & Morgan, 2009). La constatación de concepciones o visiones diferentes, así como los procesos de argumentación y contra-argumentación ante las discrepancias, puede contribuir a hacer conscientes las representaciones profundas (Marín & Cervi, 2006; Rodrigo, 1993).
 - ✓ Atender deliberadamente a las *contradicciones o disonancias que se producen entre las creencias explícitas e implícitas sobre evaluación*. Como explica R. López (2001) cuando se descubren dos creencias contrapuestas,

comienza un conflicto interno que fuerza la consciencia y el afrontamiento. Esta circunstancia y las dos que le preceden, ha llevado a investigar el potencial de los incidentes críticos vivenciados por el profesorado como estrategia formativa (Monereo, 2014b; Monereo & Badía, 2011; Monereo et al., 2013).

- ✓ Promover una formación *práctica y contextualizada*. Considerando el carácter contextual y encarnado de las creencias profundas, la formación debe centrarse en el análisis y resolución de preguntas y problemas derivados de la práctica evaluativa cotidiana, favoreciendo la integración de teoría y práctica y dando lugar a situaciones de aprendizaje significativo (Korthagen, 2010; M. Martínez, 2006; Román & Murillo, 2009). Las situaciones pueden ser reales o simuladas, pero deben conectar con la práctica haciendo del espacio formativo un escenario cotidiano (Rodrigo, 1997). Este tipo de iniciativas favorecen la superación de la brecha entre la teoría y la práctica dando lugar al replanteamiento de la relación entre ambas (Zeichner, 2010).
- La formación debe orientarse al *desarrollo de la competencia evaluativa del profesorado*. Lejos de centrarse en el análisis de contenidos evaluativos aislados (métodos, modelos o instrumentos, etc.), las iniciativas formativas deben estimular el desarrollo de la competencia docente. De acuerdo con la definición efectuada de la misma en el primer capítulo, ello supone:
 - ✓ Atender desde una perspectiva integrada y globalizadora los saberes conceptuales sobre evaluación, su puesta en práctica y los valores, principios y actitudes que deben regir dicho proceso (Trillo, 2005; Villardón, 2006).
 - ✓ Estimular la *capacidad de transferencia* a nuevas situaciones (Mateo, 2012), organizando escenarios y situaciones diversos y variados.
 - ✓ Incorporar al propio planteamiento formativo la *complejidad y la incertidumbre* (A. I. Pérez, 2010b; Santos, 2010a). Si se admite con Barnett (2005) la supercomplejidad del mundo actual, no se sostienen los planteamientos formativos simplistas, basados en supuestas relaciones lineales y estables de causalidad y efecto. Del mismo modo, si la evaluación se caracteriza por ser un proceso complejo, la formación debe evolucionar de la certeza y la simplicidad a la complejidad y la incertidumbre y preparar para ella (Santos, 2010a). “Más que una teoría educativa de la simplicidad positivista, necesitamos una teoría de la complejidad de los problemas educativos, si es que queremos resolverlos” (Batalloso, 2004, p. 46).

Los rasgos enumerados hasta el momento representan una descripción general del tipo de formación que se precisa en el contexto universitario para favorecer la mejora de los sistemas de evaluación. Olmos & Rodríguez (2010) ofrecen una visión sintética muy próxima a modo de conclusión:

Respecto a la formación docente del profesorado universitario en España, los programas de formación deberían estar basados también en competencias docentes, que le permitieran reflexionar sobre su propia práctica, en un contexto de trabajo colaborativo, donde el intercambio entre compañeros facilitará el aprendizaje de técnicas de evaluación a través de experiencias y opiniones compartidas. (p. 11).

4.2.3 Desarrollo de experiencias de innovación en materia evaluativa.

De acuerdo con las orientaciones de Marín y Cervi (2006), gestionar el cambio de las prácticas evaluativas no significa posponer la innovación hasta que las concepciones cambien, sino que la propia innovación puede representar en sí misma un estímulo en ese cambio. “Las representaciones implícitas se aprenden por la participación en contextos de práctica. Cambiar las prácticas contribuiría, pues, a modificar las representaciones, sobre todo si se estructuran procesos de reflexión de estos cambios” (p. 429).

Suscrito por V. M. López et al. (2007) o M. Martínez (2006), este planteamiento llevaría al profesorado a experimentar con sistemas de evaluación que se alejen de la perspectiva tradicional y le aporten vivencias alternativas a las acumuladas en el mismo. Por su naturaleza contextual, las creencias evaluativas profundas no se activan en situaciones novedosas (Rodrigo, 1997), por lo que la innovación representa una excelente ocasión formativa. Además de brindar nuevas vías y oportunidades para la reflexión, este tipo de experiencias puede llevar a confrontar creencias profundas y explícitas, favoreciendo el proceso de reestructuración que precisa el cambio de concepciones.

Estas experiencias de innovación nunca deben desligarse de la consideración de las premisas éticas y de los efectos que las mismas pueden acarrear (Flemming, 2003; Roach, 2003).

Con algo tan crucial como la evaluación, resulta esencial que la innovación no se trate de una manera superficial, es decir, por su propio interés. Están en juego las

carreras futuras y las vidas de los estudiantes, y por tanto el ámbito para la experimentación debe estar cuidadosamente delimitado y planificado. (Race, 2003, p. 77).

En consecuencia la introducción de innovaciones evaluativas debe considerar cautelosamente al menos las siguientes circunstancias (Capllonch & Buscà, 2012; McDowell & Sambell, 2003):

- La carga de trabajo discente que ocasiona y su impacto sobre la motivación.
- Las decisiones acreditativas que se desprenderán de la innovación puesta en práctica y su repercusión.
- Los recelos naturales ante incertidumbre y lo desconocido por parte del alumnado, así como posibles medidas para contrarrestar su efecto.
- La gestión de los probables imprevistos organizativos, asociados a la propia inexperiencia del profesorado y la dificultad de anticipación.

Situada en esta misma línea de reflexión, Bevitt (2015) advierte que resulta difícil predecir el impacto de las innovaciones evaluativas en el alumnado; en consecuencia, recomienda su implicación activa en la delimitación de los dispositivos evaluativos resultantes.

Esta tercera vía para la mejora de la praxis evaluativa guarda una estrecha conexión con la formación, alternativa previamente reseñada. El desarrollo de experiencias de innovación puede constituir en sí misma una acción formativa, en la medida que lleve parejos procesos de reflexión e investigación desde la acción y para ella (búsqueda de documentación creación de instrumentos, registro y análisis de resultados, etc.). Asimismo, puede ser consecuencia de una acción formativa previa que la motive o dar lugar a una necesidad sentida de formación.

4.2.4 Difusión de los avances, contribuciones y experiencias en materia evaluativa: investigación e intercambio.

El cambio en las representaciones evaluativas del profesorado universitario y la reducción de las disfuncionalidades de los sistemas clásicos precisa de una importante labor de difusión.

En general, el profesorado necesita tiempo, apoyo y oportunidades para el diálogo con colegas. El profesorado necesita tiempo para leer literatura profesional sobre evaluación, entablar diálogo con otros colegas en transición y tener la oportunidad de probar los nuevos procedimientos en un ambiente de apoyo y colaboración. (Serafini, 2000, p. 388).

Estrechamente ligada a la formación y la innovación, la difusión en materia evaluativa como recurso para la mejora puede llevarse a cabo en dos niveles complementarios:

- En un primer nivel se situaría la investigación formal y experta sobre evaluación del aprendizaje, que da lugar a lo que Korthagen (2010) denomina la *Teoría* con *T* mayúscula. En este sentido, son varias las voces que ponen de manifiesto la necesidad de dedicar mayor número de investigaciones a la evaluación del aprendizaje en los distintos niveles educativos, especialmente en la universidad (Ahumada, 2002, 2005; Ibarra & Rodríguez, 2010; Palacios & López, 2013). Se correspondería con el concepto de *Scholarship of teaching and learning* (Andresen, 2000; Boyer, 1990; Kreber, 2005; Shulman, 2000) que, aunque no está sujeto a una definición unívoca, puede ser concebido de forma general en los siguientes términos:

Es un proceso en el que el profesor adopta una postura reflexiva sobre su ejercicio profesional, sometiendo a examen su propio trabajo y, a continuación, comparte tanto sus experiencias como el resultado de sus reflexiones con el resto de profesionales de la enseñanza para, entre todos, avanzar en el desarrollo de una enseñanza basada en la investigación sobre la misma. (Chocarro et al., 2013, p. 7).

- En un segundo nivel, la literatura especializada insiste en la creación de redes, foros o espacios para la colaboración y el intercambio de experiencias en materia evaluativa (Caballero, 2013; B. Gargallo et al., 2010; V. M. López et al., 2007; Margalef & Álvarez, 2005; Monereo & Castelló, 2009a; Olmos & Rodríguez, 2010). Además de difundir las contribuciones derivadas del nivel anterior de investigación (la *Teoría*), estas redes ofrecen al profesorado un medio para compartir el saber y

las conclusiones construidas desde su propia experiencia, es decir, la *teoría* con *t* minúscula (Korthagen, 2010).

A un nivel o al otro, explotar la vía de avance de la difusión para la mejora de la evaluación exige superar la privacidad que acompaña a la docencia y hacer visibles las buenas prácticas para contribuir a su desarrollo profesional (Cid-Sabucedo et al., 2009).

4.2.5 Promoción y generalización de las experiencias de metaevaluación.

El establecimiento de mecanismos para potenciar y generalizar los procesos de metaevaluación representa otra vía de mejora aplicable a la evaluación del aprendizaje universitario (Durán, 2009; Monereo & Castelló, 2009a, 2009b; Weston Manor Grupo-WMG- ASKe-, 2007). Consiste en la revisión de la evaluación (Scriven, 1968) y representa la cuarta fase del proceso evaluativo, tal y como ha quedado planteado en el capítulo segundo. Se trata de un proceso para analizar y enjuiciar la praxis evaluativa en comparación con aquello que se conciba como una buena práctica (Stufflebeam, 1974/2011, 2001b). Desde esta perspectiva, Rotger (1992) y Patel (2002) afirman que la metaevaluación representa una vía eficaz para verificar y asegurar la calidad de las evaluaciones.

La contribución potencial de esta revisión a la mejora de la enseñanza y el aprendizaje se encuentra ampliamente suscrita. La literatura específica presenta la metaevaluación como un principio de calidad asociado a la praxis evaluativa (Cano et al., 2010; Boud & asociados, 2010; Brown et al., 1996). Su importancia y pertinencia, también se refleja en las contribuciones que realizan las asociaciones y agencias especializadas para favorecer su desarrollo práctico (American Evaluation Association- AEA, 2004; JCSEE, 1981, 1998, 2001, 2007; QAA, 2014).

Además de escaso, el ejercicio de la metaevaluación en el contexto formativo universitario está sujeto a ciertos tópicos que han condicionado su repercusión e impacto:

- La mayor parte de las referencias que se efectúan a la metaevaluación se centran en la revisión de la perspectiva más institucional del proceso evaluativo (normativa, programas, rendición de cuentas, política evaluativa). A ella apuntan, entre otras, las contribuciones de Stufflebeam (1999, 2001a) o la (AEA) (2004) para el desarrollo de experiencias de metaevaluación. Sin negar su necesidad e importancia, creemos que la superación de las deficiencias evaluativas descritas en el primer apartado

exige aterrizar en un nivel de reflexión más inmediato: *la evaluación del aprendizaje* en el día a día del aula. A ella se orienta el presente trabajo.

- Igual que cualquier otra evaluación, la metaevaluación puede ser formativa o sumativa (B. McDonald, 2010; Stufflebeam, 2001b), es decir, puede contribuir a los clásicos *propósitos acreditativos, así como a fines pedagógicos* vinculados a la mejora. El objetivo último de la revisión no debería ser determinar si existe o no validez, al estilo de las investigaciones experimentales, sino más bien “avivar la reflexión para generar comprensión rica y relevante sobre los programas evaluados. Esto convierte a la evaluación no tanto en una tarea de expertos cuanto en un proceso democrático de discusión” (Santos & Moreno, 2004, p. 915).
- Las escasas experiencias de metaevaluación reportadas, se asocian a la intervención de un agente externo, por lo general, con fines acreditativos. Sin embargo, concebida como recurso para la mejora de la evaluación, cabe una perspectiva más amplia y pedagógica de la revisión, entendiendo que puede ser autoguiada (*autometaevaluación*) o estar externamente dirigida (*heterometaevaluación*). Desde esta perspectiva Monereo y Castelló (2009a) señalan una vía interna y otra externa para la revisión de las evaluaciones, mientras que Stake (2006) discrimina la metaevaluación informal (autometaevaluación) de la formal (aquella en la que se cuenta con supervisión externa).
- Las modalidades de revisión señaladas centran la atención en el profesorado y en agentes externos, pero su *participación* puede ser complementada con la de otras fuentes. La visión que proporciona el alumnado de su experiencia en el sistema de evaluación puede representar una potente herramienta formativa para el ajuste de deficiencias y el diseño de sistemas más efectivos y eficaces (Batalloso, 2004; Flemming, 2003; Leal, 2004; B. McDonald, 2010). Dependiendo de los aspectos concretos en los que repare la metaevaluación, colegas y superiores también pueden ofrecer un testimonio valioso y pertinente (Capllonch et al., 2009). Boud y asociados (2010), por su parte, sugieren la participación de alumnado egresado y empleadores en el proceso de revisión, argumentando que la verdadera contribución de la evaluación a la formación y el aprendizaje no puede valorarse hasta la incursión en el mundo profesional.
- A pesar de integrar en su calificativo el prefijo -Meta-, la metaevaluación *no tiene por qué llegar finalizada la evaluación*, sino que puede tener lugar durante su *planificación y en la implementación o puesta en práctica* (Santos & Moreno, 2004). Desde esta perspectiva, Santos (2003) discrimina entre la revisión proactiva

(realizada en la fase de planificación), la interactiva (en el transcurso de la aplicación) o la retroactiva (finalizada la implementación).

- Las directrices y estándares que rigen los procesos de metaevaluación se centran habitualmente en las dimensiones técnica y ética del proceso. Así se constata, por ejemplo, en las propuestas del JCSEE (1981, 1998, 2001, 2007), Stufflebeam (1999) o de la African Evaluation Association (AfrEA, 2007). Sin embargo, situados en la formación universitaria, la revisión de la evaluación debería detenerse en la valoración del efecto que ocasiona en el aprendizaje del alumnado y su mejora (Hargreaves, 2007; Mateo, 2000). Así lo exige la adopción de la evaluación para el aprendizaje como marco de referencia conceptual.
- Las pocas experiencias de metaevaluación que se producen, por lo general, revierten de forma exclusiva en el foco exacto en el que han sido efectuadas. Sin embargo, las conclusiones extraídas de un proceso de metaevaluación pueden ser útiles para estimular el aprendizaje y mejora en otros foros, por lo que debería de impulsarse su *puesta en común y transferencia* (Moreno & Santos, 2004; Santos, 2003). Este planteamiento remite a la cuarta línea de mejora contemplada, relativa a la difusión.

En cualquiera de las modalidades sugeridas, la metaevaluación debe partir de la actitud crítica y abierta de las/os participantes (Ahumada, 2005; Porto, 2005a) y no ser percibida como un riesgo, una amenaza o un mecanismo de control y supervisión (Casanova, 1992). Para ello, debe transcurrir en condiciones cuidadosamente delimitadas que configuren un clima seguro, propicio para el aprendizaje y el desarrollo profesional. Dependiendo de los propósitos y objetos concretos a los que haga referencia la revisión de la evaluación, ésta puede recurrir a marcos de referencia complementarios para favorecer que así sea:

- La enumeración de principios de referencia que ofrece la literatura para el desarrollo de cualquier proceso de evaluación, incluidas las metaevaluaciones. Sirvan a modo de ejemplo los principios sintetizados por Care (1978), Rotger (1992), Ortiz (2006) o Monereo y Castelló (2009a), entre otros.
- El marco general de referencia que ofrecen las propuestas estandarizadas para el desarrollo de las evaluaciones (JCSEE, 1981, 1988, 1998, 2007; AfrEA, 2007).
- Las orientaciones específicas para el desarrollo de experiencias de metaevaluación (Stufflebeam, 1999, 2001a).

- Los principios rectores propuestos para la práctica profesional y responsable de las/os evaluadores (AEA, 2004).

5. Los sistemas tradicionales de evaluación del aprendizaje en la formación inicial de maestras/os de Educación Infantil y Primaria

La radiografía de las prácticas evaluativas tradicionales en la universidad revela todo un conglomerado de síntomas, deficiencias y contradicciones según ha sido advertido. Tal circunstancia resulta especialmente alarmante en la formación de futuras/os maestras/os, donde además de una herramienta pedagógica y acreditadora, la evaluación del aprendizaje representa una de las competencias específicas a desarrollar entre el alumnado (Bilbao, 2010, 2011a). La capacidad de evaluación y autoevaluación está reconocida como competencia genérica o transversal en cualquier ámbito profesional, en la medida que permite distinguir lo correcto y lo inadmisible en la propia actuación o en la de otros (Boud, 2000; Boud & Falchikov, 2006; Dochy et al., 1999; Ljungman & Silen, 2008; Villa, 2008; Vu & Dall'Alba, 2007). Pero aún más allá, en las titulaciones vinculadas a la formación básica de maestras/os la evaluación del aprendizaje aparece recogida como objetivo y competencia específica a desarrollar. Así se hace constar en las órdenes 3854/2007 y 3857/2007 que establecen los requisitos para la verificación de los títulos de maestra/o en Educación Infantil y Educación Primaria, respectivamente²⁷.

Sin embargo, la especial significación de la evaluación en estas titulaciones no se traduce en una praxis evaluativa de mayor calidad según los diagnósticos consultados. *En casa del herrero, cuchillo de palo*, se lamenta V. M. López (2004) y es que, en las facultades de educación y escuelas de magisterio, la evaluación del aprendizaje no muestra una cara muy diferente de la descrita a nivel general en el primer apartado. Así lo refleja la tabla 25, que ofrece una síntesis de las características evaluativas reportadas en estudios específicamente contextualizados en la formación de maestras/os o profesorado:

²⁷ El contenido específico de dichas regulaciones será sometido a análisis en el capítulo próximo.

Tabla 25

Rasgos característicos de la evaluación del aprendizaje en el contexto específico de la formación básica de maestras/os

EVALUACIÓN DEL APRENDIZAJE EN LA FORMACIÓN BÁSICA DE MAESTRAS/OS Resumen de los rasgos reportados en la investigación específica	
FINALIDAD	• Principalmente acreditativa, centrada en la calificación. • Unido a lo anterior, la evaluación se emplea como mecanismo de control fundamentalmente.
MOMENTO	• La evaluación es prioritariamente final y no se considera plenamente integrada en el proceso formativo. • Apenas se llevan a cabo evaluaciones iniciales.
INSTRUMENTOS Y TAREAS	• Prevalece la evaluación cuantitativa, en ocasiones realizada mediante instrumentos inadecuados. • Por lo general las tareas evaluativas son descontextualizadas, basadas en la reproducción de saberes declarativos. • También en la formación inicial de maestras/os el examen representa el método de evaluación por excelencia.
AGENTES	• El profesorado es el agente evaluador por excelencia. • La participación del alumnado en la evaluación es infrecuente y limitada. Por lo general se restringe a tareas de corrección desarrolladas durante la puesta en práctica.
REVISIÓN	• Las experiencias de metaevaluación son infrecuentes.

Nota: Síntesis elaborada a partir de varias aportaciones (Arribas et al., 2010; C. Brew et al., 2009; Ferguson, 2011; Gutiérrez et al., 2011; Gutiérrez et al., 2013; Hamodi & López, 2012; V. M. López, Manrique et al., 2011; V. M. López & Palacios, 2012; L. F. Martínez et al., 2012; Palacios & López, 2013; M. L. Pérez et al., 2000; Ruiz-Gallardo et al., 2013; Smitter, 2006; Smitter & Siso, 2007).

Pero de las disfuncionalidades constatadas a nivel general, y replicadas en la situación particular de la formación básica de los maestras/os, la más alarmante de todas es la grave falta de coherencia entre el discurso teórico que se sostiene sobre evaluación y su traslación práctica (Hamodi & López, 2012; V. M. López, Manrique et al., 2011; Palacios & López, 2013; Smitter & Siso, 2007; Tonucci, 2010). “Es frecuente el divorcio entre lo que se explica y lo que se hace. La evaluación que realizan los alumnos contradice las teorías que aparecen en los libros sobre evaluación que los alumnos estudian para realizar sus exámenes” (Santos, 1999, p. 380).

Las incoherencias constituyen una dificultad en la formación general del profesorado, por restar impacto y credibilidad y obstaculizar el proceso de sinergia (J. Contreras, 2010; Perrenoud, 2010). En el terreno que nos ocupa la más directa de las consecuencias es la negación de los principios de la evaluación (Santos, 2009). Apoyados en el bagaje teórico que les proporciona su iniciada trayectoria formativa, las/os estudiantes o futuras/os maestras/s viven y señalan la experiencia de la contradicción. De acuerdo con

V. M. López (2004) el aprendizaje colateral que deja esta vivencia es que en educación se puede ser incoherente, cuando no hipócrita.

Es aún más grave, y muestra de una considerable dejadez profesional, cuando ya se sabe, desde hace varias décadas, que el profesorado no tiende a hacer lo que durante su formación inicial le contaron que había que hacer, sino lo que hicieron con él como alumno. (p. 226).

Los aprendizajes profundos sobre la evaluación se inician en los niveles escolares (M. A. Calatayud, 2004) y se desarrollan a lo largo de toda la trayectoria académica a partir de las experiencias acumuladas. En el caso concreto de las/os maestras/os, una parte importante de dicha trayectoria tiene lugar en las facultades de educación y escuelas de magisterio. Consideramos que en ese punto se produce un círculo vicioso que aleja las prácticas evaluativas de la mejora y que recuerda a la popular analogía de la *pescadilla que se muerde la cola* (ver figura 14).



Figura 14. Ciclo de reproducción de las prácticas evaluativas disfuncionales por medio de la formación básica de maestras/os.

La generación de este ciclo de reproducción responde a la concatenación de dos circunstancias:

- Por una parte, replicando su vivencia estudiantil, e impulsado por las mismas circunstancias estructurales y profundas que confluyen en otras titulaciones, el profesorado de los grados de formación básica de maestras/os emplea dispositivos evaluativos disfuncionales.
- Por la otra, el alumnado matriculado (futuras/os maestras/os y evaluadoras/es) va configurando representaciones evaluativas en consonancia con dicha vivencia. Como ya ha sido aclarado, estas concepciones profundas construidas desde la experiencia condicionan poderosamente la actuación evaluativa futura, predisponiendo su reproducción.

Santos (2009) emplea la metáfora de las *almas tatuadas* para referirse a quienes, convertidos en evaluadoras/es, incurren en los mismos errores e injusticias que como estudiantes lamentaron. Uno de los principales indicios que sugiere el buen funcionamiento de este ciclo de reproducción es que la cultura evaluativa escolar, ámbito de intervención profesional de maestras/os, muestra muchos de los rasgos patológicos diagnosticados en la universitaria²⁸.

La persistencia o perpetuación de las prácticas evaluativas disfuncionales, desde los niveles educativos escolares y hasta los universitarios, parece guardar una estrecha relación con la naturaleza de las convicciones evaluativas profundas y con el enfoque tradicionalmente adoptado en la formación básica de maestras/os. El alumnado de magisterio concluye sus estudios con concepciones sobre la enseñanza, el aprendizaje y la evaluación muy pobres desde la perspectiva constructivista (Schommer et al., 2012). El hecho de que la evaluación no llegue a verse como un recurso didáctico privilegiado después de años de formación, puede deberse a dos causas particularmente:

- Por una parte, la capacitación evaluativa explícita se encuentra escasamente representada dentro de los planes de estudio para la formación básica de maestras/os (Mertler, 2009; Stiggins, 1993/2012, 2002; Stiggins & Conklin, 1998). Si bien es cierto que los aprendizajes declarativos sobre la evaluación no garantizan la modificación de las representaciones profundas que guían la práctica, esta apreciación sugiere una escasa reflexión general sobre el fenómeno evaluativo en estas titulaciones.

²⁸ Según las descripciones planteadas por Román y Murillo (2009), Santos (2003) o D. Salinas (2002).

- Por la otra, las concepciones evaluativas con las que accede el alumnado a la formación básica, tras largos años de experiencia en el sistema educativo, no son combatidas de acuerdo con las orientaciones ofrecidas en la tabla 24. Por una parte, son frecuentes las *prácticas ciegas* (Pozo et al., 2006), es decir, aquellas iniciativas formativas que confían en la mera exposición y repetición de modelos teóricos para el cambio de representaciones evaluativas. Por la otra, tampoco se considera el impacto que tiene en su persistencia el mensaje implícito que emite la práctica evaluativa directa del propio profesorado universitario, a menudo, poco consistente con el discurso teórico (Palacios & López, 2013).

En síntesis, consideramos que existen razones poderosas para requerir del profesorado universitario la oferta de un modelo de evaluación del aprendizaje que refuerce desde la práctica los principios teóricos difundidos y el tipo de evaluación que pretenden promoverse entre las/os futuras/os maestras/os. Igual que ha sido sugerido para la formación universitaria en general, el avance debe realizarse atendiendo a los factores estructurales y profundos que condicionan la praxis evaluativa.

6. El capítulo a grandes rasgos

En el capítulo que concluye se ha llevado a cabo una revisión de los sistemas de evaluación del aprendizaje que se emplean en el sistema universitario español en general, y en las titulaciones vinculadas a la formación básica de maestras/os en particular. El recorrido efectuado nos ha conducido a las siguientes conclusiones:

- Los sistemas que se emplean en el contexto universitario para la evaluación del aprendizaje de las/os estudiantes son **disfuncionales**. Manifiestan síntomas importantes de deficiencia y encierran en sí mismos graves contradicciones e incoherencias, advertidas desde diferentes puntos de vista.
- Esta disfuncionalidad provoca **efectos negativos** tanto en docentes, como en discentes.

✓ En cuanto a las/os *discentes*...

- ⇒ Sufren las consecuencias de procesos de selección en los que sólo se consideran sus calificaciones, aunque éstas hayan sido adjudicadas en el marco de sistemas de evaluación deficientes.
- ⇒ Desarrollan sentimientos y actitudes negativas hacia la evaluación desde edades muy tempranas.
- ⇒ Aunque superen el módulo/asignatura, no aprenden o acumulan aprendizajes superficiales.
- ⇒ Construyen representaciones o concepciones profundas sobre la evaluación de acuerdo a la experiencia vivenciada y actúan de acuerdo a ellas en las situaciones formativas que se les plantean en su trayectoria académica.

✓ El *profesorado*, por su parte...

- ⇒ Desarrolla una visión negativa de la evaluación, en paralelo a la que manifiestan las/os estudiantes.
- ⇒ Ve reducido el efecto de su enseñanza, que pierde potencialidad didáctica por el uso ineficaz e ineficiente de la evaluación.
- ⇒ Unido a lo anterior, reduce sus creencias de autoeficacia docente.
- ⇒ Ante la generalización de las disfuncionalidades, incurre en una actitud de apatía o inmovilismo profesional, caracterizado por la falta de cuestionamiento.

● Las deficiencias detectadas y su persistencia en el tiempo se asocian a la **incidencia de una serie de factores de naturaleza diversa. Junto al influjo de las condiciones estructurales, se aprecia la incidencia de otra serie de circunstancias más profundas**, que exceden la dimensión material y práctica de los medios y recursos:

✓ Las *condiciones estructurales* que en mayor medida condicionan las disfuncionalidades detectadas en los sistemas evaluativos y su persistencia son:

- ⇒ La incidencia de normas y reglamentos ideados desde la perspectiva administrativa, en vez de desde la lógica didáctico-pedagógica.

- ⇒ El elevado número de estudiantes con el que se trabaja habitualmente en la universidad y la creciente diversidad en los perfiles de acceso.
- ⇒ La incidencia de dificultades vinculadas al estatus socio-laboral del profesorado universitario (sobrecarga, precariedad, devaluación de la docencia y presión investigadora, etc.).
- ⇒ La sobrecarga que generan en docentes y discentes los sistemas de evaluación más próximos a planteamientos pedagógicos (evaluación continua, formativa, etc.) y las resistencias consiguientes.
- ⇒ Las presiones sociales por la calificación, que obstaculizan procesos de innovación evaluativa por medio de recelos y resistencias.
- ⇒ La confusión y ambigüedad terminológica que se produce en el terreno de la evaluación, que genera inseguridades y dificulta procesos de intercambio y trabajo compartido de cara a la mejora.
- ✓ En cuanto a los *factores de naturaleza profunda*, aquellos que mejor explican la situación de disfuncionalidad y su mantenimiento en el tiempo son:
 - ⇒ Una falta manifiesta de cultura evaluativa que afecta a los distintos niveles del sistema educativo.
 - ⇒ Una insuficiente e inadecuada capacitación evaluativa entre el profesorado universitario que, por lo general, accede a la docencia sin formación específica en este terreno.
 - ⇒ Unas concepciones negativas sobre evaluación, profundamente arraigadas y extendidas entre alumnado y profesorado universitario, a raíz de las vivencias acumuladas por unos y otros en largos años de trayectoria académica.
- Para reducir la incidencia de las disfuncionalidades detectadas e introducir cambios significativos en las prácticas y las representaciones profundas sostenidas sobre evaluación, es preciso **intervenir sobre ambos tipos de causa**:
 - ✓ Es necesario atender a la parte estructural, procurando la revisión y redefinición de las condiciones laborales en las que se ejerce la docencia y del contexto administrativo y organizativo en el que transcurre la evaluación.

- ✓ Muy en especial, exige emprender iniciativas para afectar a las condiciones profundas que favorecen la disfuncionalidad evaluativa actual. Para ello, resulta prioritario:
 - ⇒ Promover entre el profesorado una actitud de humildad, revisión y autocrítica con relación al fenómeno evaluativo.
 - ⇒ Mejorar la formación en evaluación del profesorado universitario cuantitativa y cualitativamente.
 - ⇒ Desarrollar experiencias de innovación que generen nuevas vivencias evaluativas y contribuyan a cuestionar el patrón evaluativo tradicional.
 - ⇒ Fomentar la investigación y el intercambio de experiencias sobre la evaluación del aprendizaje en el contexto universitario.
 - ⇒ Promocionar y generalizar las experiencias de metaevaluación desde la perspectiva pedagógica del aprendizaje y la mejora profesional.
- El planteamiento general de la mejora de la evaluación del aprendizaje universitario adopta una **significación especial en aquellas titulaciones vinculadas a la formación básica de maestras/os**. El efecto de las disfuncionalidades evaluativas constatadas es mucho más grave en este caso, donde la evaluación figura entre las competencias u objetivos de aprendizaje a desarrollar por parte de las/os estudiantes. Ejercida ante futuras/os evaluadoras/es, el uso que lleve a cabo el profesorado universitario de la evaluación, como recursos didáctico y acreditativo, ofrece un mensaje adicional y modela los patrones evaluativos que serán reproducidos por las nuevas generaciones docentes, tanto si son positivos como negativos.

CAPÍTULO 4.

LA EVALUACIÓN DEL APRENDIZAJE EN LA UNIVERSIDAD DEL SIGLO XXI

La construcción de un modelo que pormenorice la competencia evaluativa del profesorado dedicado a la formación básica de maestras/os de Educación Infantil y Primaria reposa necesariamente sobre una determinada forma de concebir e interpretar la naturaleza de la evaluación del aprendizaje universitario. Admitido que coexisten visiones y concepciones diferentes en torno a este fenómeno, el nuevo capítulo tiene por objeto detallar la particular teoría a la que se orienta la presente investigación.

Las próximas páginas enumeran los rasgos y características de la teoría de la evaluación que pretende consolidarse con este trabajo. El resultado final de la deliberación se muestra estructurado en el modelo denominado ITC (Integrado Teórico Complejo), que enumera subprocesos, criterios e indicadores característicos de la praxis evaluativa deseable desde nuestra particular perspectiva. Dicho modelo, constituye el *debería ser* o la teoría de la evaluación del aprendizaje a la que se orienta la propuesta de competencia evaluativa docente expuesta en el capítulo siguiente.

El capítulo se estructura en cuatro grandes apartados:

- El primer apartado comienza con la revisión de los principales referentes de los que se nutre la visión o teoría de la evaluación propuesta. En concreto, detalla las aportaciones recogidas de tres grandes fuentes de influencia: el paradigma y la conceptualización teórica de partida, las particularidades del contexto evaluativo al que se orienta la investigación y los antecedentes que han tratado de sistematizar rasgos, principios o estándares para la práctica evaluativa deseable en la universidad.
- Considerados en su conjunto, dichos referentes dan lugar a la estructuración del modelo ITC, teoría de la evaluación asumida como referencia en la presente investigación. El segundo apartado recorre sus principales rasgos y justifica la ordenación general de la propuesta.

- El tercer apartado expone el desarrollo del modelo completo; por cada uno de los subprocesos, presenta en forma de ficha los criterios e indicadores articulados para su transferencia práctica.
- El capítulo concluye con una síntesis global de las decisiones y conclusiones más relevantes adoptadas a lo largo del mismo.

1. Referentes para la definición de una teoría de la evaluación del aprendizaje de las/os estudiantes en la universidad del siglo XXI

En torno al fenómeno evaluativo coexisten teorías o perspectivas parciales. Su credibilidad, rigor y relevancia está determinada por la solidez de las bases sobre las que reposa su fundamentación y justificación. Para seleccionar los principios y consideraciones sobre las que construir el modelo ITC de evaluación, se ha recurrido a tres grandes referentes o fuentes de influencia (ver figura 15):

- La posición teórica desde la que se interpreta el sentido y la relevancia del proceso evaluativo en el presente trabajo.
- Las condiciones inmediatas que configuran su contexto de aplicación potencial. En el marco de la presente investigación, ello supone la revisión de tres factores relevantes: las implicaciones evaluativas sujetas al EEES, los referentes normativos que regulan la evaluación del aprendizaje discente en la universidad y el especial papel que adopta la evaluación como competencia específica a desarrollar en la formación básica de maestras/os de Educación Infantil y Primaria.
- Los antecedentes que han tratado de definir un modelo de evaluación del aprendizaje hacia el que aproximar la formación universitaria. Esta tercera fuente reúne aportaciones de distinto tipo y desigual nivel de desarrollo (manifiestos y formulación de principios evaluativos de referencia, propuestas más estructuradas que delimitan criterios y pautas de actuación, así como propuestas estandarizadas).



Figura 15. Referencias para la construcción de una teoría de la evaluación (modelo ITC).

La consideración conjunta de estos referentes conduce a una determinada forma de entender la evaluación del aprendizaje discente en el contexto universitario específico acotado. Los próximos subapartados detallan de forma sintética la influencia recibida de cada uno de ellos. El modelo concreto que desgrana tal visión es el **modelo ITC**, presentado en el segundo apartado del capítulo.

1.1 Posición teórica de partida: la evaluación como indagación y como recurso estratégico para la promoción del aprendizaje.

A lo largo de su evolución histórica, la evaluación del aprendizaje ha generado en torno a sí interpretaciones diferentes que la han calificado como medida, procedimiento, control de la calidad o indagación (Falchikov, 2005; Serafini, 2000). Centrada en este último paradigma, la investigación en curso se nutre de las diversas conceptualizaciones articuladas en su desarrollo, todas ellas enmarcadas dentro del enfoque de la evaluación alternativa. Según ha sido justificado en el capítulo segundo, como principal referencia

adopta la perspectiva de la *evaluación para el aprendizaje*, en los términos que serán precisados en las próximas líneas.

La situación descrita en el capítulo anterior lleva a colocar la evaluación universitaria en el polo de la *evaluación del aprendizaje*, caracterizado por reducir el proceso a la calificación y la rendición de cuentas (Arter, 2009; Banta, 2007; Harlen, 2003). La *evaluación para el aprendizaje* surge en reacción al mismo, en el contexto escolar del Reino Unido y de la mano del Assessment Reform Group (ARG, 1999). Su extensión progresiva al ámbito universitario tiene lugar gracias a las aportaciones de S. Chappuis & Stiggins (2000), Stiggins (2002, 2005a) o Carless (2002, 2007), dentro de una extensa lista de referencias que contribuyen a su consolidación.

“La evaluación para el aprendizaje es un intento consciente de hacer de la evaluación un elemento productivo del proceso de aprendizaje” (Stobart, 2010, p. 168). Más allá de la acreditación de resultados y la rendición de cuentas, propone para ella una misión pedagógica centrada en la estimulación deliberada y el apoyo del aprendizaje (Carless, 2002, 2007; J. Chappuis, 2014; Clark, 2012; Shepard, 2000). Denotando su clara adscripción al paradigma de la indagación, el ARG (2002) la define como “el proceso de buscar e interpretar evidencias para que estudiantes y profesores conozcan dónde se encuentra el alumno en relación a su aprendizaje, dónde necesita estar y cuál es el modo mejor de llegar ahí” (p. 2).

“La evaluación para el aprendizaje convierte la evaluación del día a día en un proceso de enseñanza y aprendizaje que mejora el aprendizaje de los estudiantes, en vez de monitorearlo solamente” (Stiggins, 2007, p. 22). Así concebida, la adecuación y eficacia de la evaluación se determina por su contribución pedagógica, en función de la capacidad que muestre para promover en el alumnado aprendizajes de calidad (Gielen, Dochy, & Dierick, 2003; Hargreaves, 2007; Monereo & Castelló, 2009a). Como resume A. Fernández (2010),

el cambio es, ante todo cultural, ya que supone una visión diferente sobre la naturaleza del aprendizaje y del papel de la evaluación. En esta nueva concepción la evaluación se sitúa en el centro del proceso educativo y su uso se justifica en tanto en cuanto regula la calidad de los aprendizajes y, por ende, la calidad de la docencia universitaria. (p. 13).

Frente a quienes consideran que la evaluación formativa es una denominación más para designar la *evaluación para el aprendizaje* (Hargreaves, 2007; Harlen, 2005, 2006), en

la presente investigación sostenemos que no son la misma cosa, siguiendo la referencia de una larga lista de autores (Carless, 2007; Clark, 2012; Stiggins, 2005a, 2005b; Stiggins & Chappuis, 2008; Wiliam, 2011). Desde nuestra perspectiva, la evaluación para el aprendizaje ofrece un marco de interpretación más amplio, en el que pueden integrarse las distintas funciones evaluativas, tanto las pedagógicas como las sociales (ver figura 16):

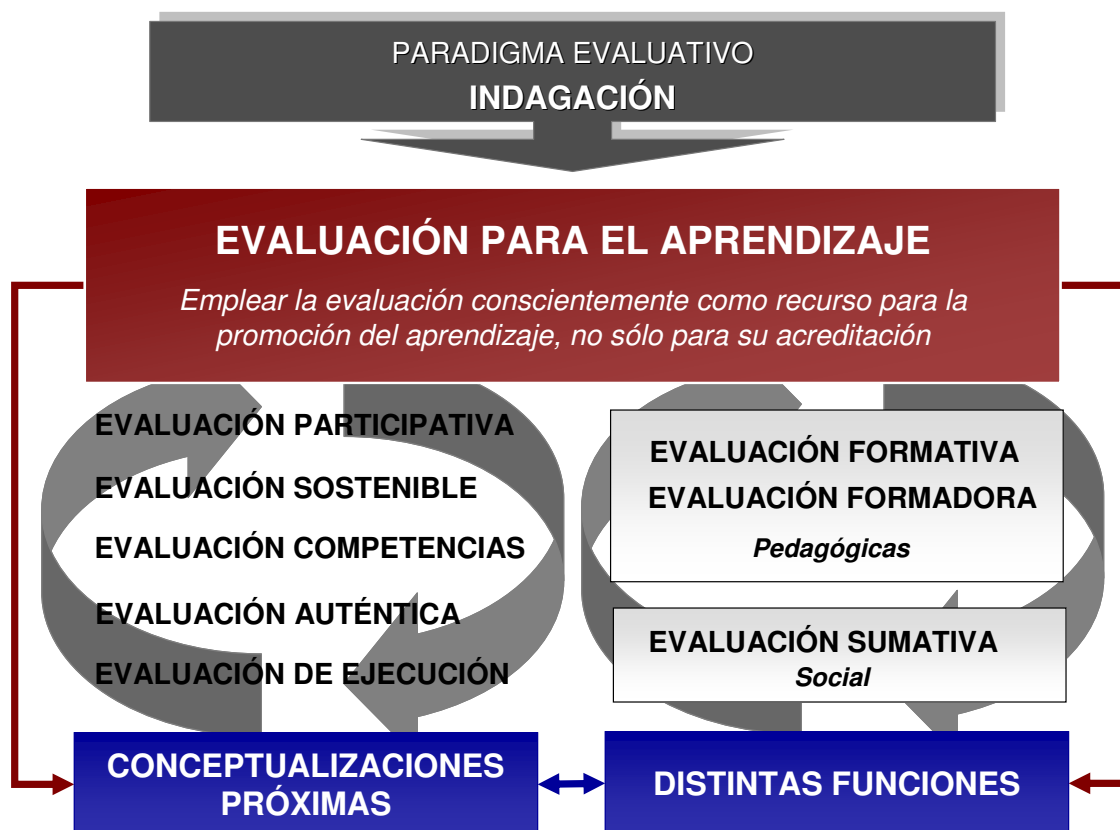


Figura 16. Delimitación conceptual de la evaluación para el aprendizaje en el marco de la investigación en curso.

Siempre y cuando el propósito prioritario de su confluencia sea la promoción del aprendizaje y no sólo su medición, esta visión de la evaluación permite conjugar las funciones formativa, formadora y sumativa. El uso pedagógico de la evaluación es prioritario desde esta posición, pero no exclusivo. Hablar de evaluación sumativa es hablar de evaluación *del* aprendizaje, pero ésta puede adoptar una orientación pedagógica según se emplee:

- Por una parte es preciso considerar que se puede hacer un uso formativo de las conclusiones de una evaluación sumativa (Black & Wiliam, 2009; Winniger, 2005). Aunque los resultados puestos de manifiesto tengan valor acreditativo, pueden proporcionar a estudiantes y profesorado datos relevantes sobre los que emprender

acciones de mejora, que refuercen y contribuyan el aprendizaje (S. Chappuis & Chappuis, 2007).

- Por la otra, el poder movilizador de las calificaciones, como fuente de motivación externa, también puede ser puesto al servicio de fines pedagógicos (Race, 2003). Empleada estratégicamente y de forma sensible, la presión acreditativa puede ayudar a regular ciclos de trabajo y esfuerzo en momentos clave del proceso de aprendizaje.

Carless, Joughin, Liu y asociados (2006) participan de esta visión y entienden que la evaluación para el aprendizaje no está cerrada a la función acreditativa. Evidentemente, para ello la evaluación sumativa debe tener lugar en condiciones que no enturbien los propósitos pedagógicos del sistema (sólo en el momento adecuado, nunca de manera exclusiva, planteando demandas coherentes con los aprendizajes previstos, sin desligarse del proceso, etc.).

La amplitud de la evaluación para el aprendizaje así concebida, no sólo permite la integración de las diversas funciones evaluativas, sino también la convergencia de las distintas interpretaciones que surgen desde el paradigma de la evaluación como indagación. Aunque cada una de ellas tenga sus propios matices, la postura adoptada ofrece un marco de referencia amplio en el que tienen cabida el resto de conceptualizaciones representadas en la figura 16 (evaluación sostenible, participativa, basada en competencias, auténtica y del desempeño). Desde nuestro punto de vista, la evaluación para el aprendizaje representa la esencia y el denominador común de todas ellas; determina la vocación esencialmente pedagógica de la evaluación y las diversas interpretaciones aportan matices específicos desde donde ejercerla.

La evaluación para el aprendizaje representa un movimiento bien diferenciado en el Reino Unido y su desarrollo internacional es cada vez mayor, pero no se ha organizado como teoría independiente (Stobart, 2010). Aun así, en las contribuciones que planten sus partidarias/os se advierten una serie de premisas o supuestos teóricos comunes. Para sintetizar los más característicos y relevantes, hemos seleccionado cuatro referencias clave que han atendido a la enumeración de rasgos y condiciones básicas para ejercer la evaluación para el aprendizaje. Con amplio reflejo en la literatura, las cuatro proceden del contexto específico del Reino Unido; aunque se dirigen al ámbito escolar prioritariamente, la amplitud de las premisas que incorporan permite su fácil extrapolación al espacio universitario:

1. *Principios de la evaluación para el aprendizaje* (ARG, 2002). Una primera referencia ineludible es la que ofrece el Assessment Reform Group (2002), partiendo de la revisión de una primera propuesta elaborada tres años antes (ARG, 1999). Representa la aproximación más sistemática y ordenada a esta conceptualización evaluativa. En ella se enumeran diez principios generales para guiar la evaluación para el aprendizaje.
2. *Rasgos de la evaluación para el aprendizaje* (Stiggins, 2005a, 2005b). La segunda referencia consiste en un trabajo de conceptualización que delimita la diferencia entre la evaluación para el aprendizaje y la evaluación formativa. Sienta así las bases para definir esta conceptualización evaluativa y ofrece una primera aproximación a las claves y estrategias que desarrollan Stiggins, Arter, Chappuis y Chappuis (2006) en el próximo referente.
3. *Cinco claves y siete estrategias para la evaluación para el aprendizaje* (Stiggins et al., 2006²⁹). La contribución de estos autores es ampliamente referida en la literatura y objeto de diversos desarrollos y profundizaciones (Arter, 2009; J. Chappuis, 2009; S. Chappuis, Chappuis, & Stiggins, 2009). Por un lado, formulan cinco claves fundamentales para trasladar a la práctica la evaluación para el aprendizaje, con una serie de indicadores para orientar la tarea docente. Por el otro, proponen para su consecución siete estrategias, ordenadas de acuerdo a una estructura de tres niveles. Esta última, es deudora de las contribuciones iniciales de Atkin, Black y Coffey (2001) sobre los niveles de retroalimentación evaluativa y su repercusión en el aprendizaje.
4. *Principios de la evaluación orientada al aprendizaje* (Carless, 2007). Esta cuarta referencia es la única referencia que no se circunscribe específicamente al ámbito escolar. En ella se define la evaluación orientada al aprendizaje y se establecen tres principios clave para su desarrollo. En coherencia con la conceptualización realizada en el capítulo segundo, el matiz de la *orientación* al que se inclina esta propuesta lleva a destacar especialmente la participación discente en la evaluación y su efecto prospectivo en el desarrollo de la capacidad de aprendizaje.

El análisis y síntesis de las aportaciones desprendidas de estas fuentes ha dado lugar a la identificación de cinco núcleos centrales, indispensables para ejercer la

²⁹ La segunda edición de la propuesta la firman J. Chappuis et al. (2012) y así ha sido referida en capítulos anteriores; en este caso se recurre a la referencia inicial por preservar la visión cronológica del recorrido.

evaluación para el aprendizaje (ver la tabla 26). Entendemos, en consecuencia, que la configuración de un modelo de evaluación desde esta perspectiva teórica no puede obviar la consideración de los mismos:

Tabla 26
Principales rasgos de la evaluación para el aprendizaje

EVALUACIÓN PARA EL APRENDIZAJE: RASGOS	REFERENCIAS			
	1	2	3	4
CENTRADA EN EL APRENDIZAJE				
<i>La evaluación está plenamente integrada en el proceso de enseñanza-aprendizaje.</i>	X	X		
<i>La evaluación está enfocada a favorecer el proceso de aprendizaje deliberadamente.</i>	X		X	
<i>Las tareas de evaluación son tareas de aprendizaje.</i>				X
IMPACTO EMOCIONAL Y MOTIVACIÓN				
<i>La evaluación es ejercida considerando el impacto emocional y motivacional que ocasiona en el alumnado y su aprendizaje.</i>	X		X	
INTERIORIZACIÓN DE CRITERIOS Y CONDICIONES				
<i>La evaluación reposa sobre una definición clara y explícita de los logros de aprendizaje que pretende constatar e incorpora mecanismos para que el alumnado los interiorice.</i>	X	X	X	X
GUÍA CONSTRUCTIVA PARA EL APRENDIZAJE				
<i>La evaluación favorece la generación de feedback.</i>	X	X	X	X
<i>La evaluación proporciona pautas para la proacción.</i>	X	X	X	X
IMPLICACIÓN DEL ALUMNADO EN LA EVALUACIÓN				
<i>La evaluación incorpora deliberadamente diversas modalidades de participación discente para contribuir a sus propósitos.</i>	X	X	X	X

Nota: Síntesis efectuada a partir de las cuatro propuestas presentadas: (1) ARG (2002); (2) Stiggins (2005a, 2005b); (3) Stiggins et al. (2006) y J. Chappuis et al. (2012); (4) Carless (2007).

Tan sólo se señalan los aspectos referidos de forma explícita en las propuestas; dada la estrecha interconexión de las cinco condiciones, la interpretación no literal de las referencias llevaría a concluir que están presentes en todas ellas.

La pertinencia de la síntesis presentada encuentra respaldo en otras aproximaciones que se han efectuado a esta conceptualización evaluativa desde la literatura:

- Carless, Joughin y Mok (2006), Ibarra, Rodríguez y Gómez (2012), Palacios y López (2012) e Ibarra y Rodríguez (2010) señalan como rasgos fundamentales de la evaluación para el aprendizaje tres de las cinco condiciones contempladas: la delimitación de tareas de evaluación, la implicación del alumnado en el proceso de

evaluación y la comisión de feedback y pautas de mejora a partir de los resultados de aprendizaje constatados³⁰.

- Knight (2005), por su parte, apunta a otras tres claves que también se encuentran en la recopilación propuesta: la adecuación de las tareas, la formulación de criterios y condiciones comprensibles y la provisión de feedback.

1.2 La evaluación del aprendizaje en el contexto universitario de referencia.

Para ofrecer una respuesta adecuada a las condiciones y circunstancias del entorno de aplicación previsto, la configuración de una teoría evaluativa debería partir del análisis del contexto de aplicación al que se orienta. La presente investigación proyecta la definición de un modelo de evaluación del aprendizaje para los grados asociados a la formación de maestras/os de Educación Infantil o Primaria. Como reflejaba la figura 15, en este espacio específico es preciso considerar, al menos, tres influencias contextuales relevantes:

- Las exigencias evaluativas derivadas del EEES, como marco general de referencia para guiar el proceso de evaluación en las universidades europeas.
- Los referentes normativos que regulan la evaluación del aprendizaje en el contexto universitario español desde la perspectiva legal y administrativa.
- El propio currículo de las titulaciones reseñadas, que hace de la evaluación un resultado de aprendizaje, además de un medio didáctico y acreditativo.

Los tres próximos apartados detallan las aportaciones desprendidas de cada una de estas referencias para orientar la definición de una teoría de la evaluación adecuada a su contexto. La exposición se efectúa de forma sintética resumiendo en cada caso las conclusiones extraídas.

³⁰ Dicha selección coincide plenamente con la apuesta de trabajo del grupo EVALFOR de la Universidad de Cádiz (<http://sej509.uca.es/proyectos>), comprometido en la promoción de prácticas evaluativas formativas. Desarrolla en consecuencia proyectos que tienen que ver con la participación (EVALPART, Proyecto Alfa), con la proacción (PROALeval) o con el desarrollo de tareas y prácticas evaluativas para el desarrollo de competencias (INEVALCO, Re-Evalúa).

1.2.1 Necesidades y exigencias del EEES con respecto a la evaluación del aprendizaje.

Muchos de los cambios que precisa la consolidación del EEES tienen que ver de forma directa con la evaluación y así se refleja en las directrices generales que proporciona la ENQA (European Association for Quality Assurance, 2009). El proceso de convergencia lleva pareja una determinada visión del fenómeno evaluativo con influencia en la evaluación de las instituciones, del personal y, particularmente, del aprendizaje discente. Como refleja la tabla 27, los cambios evaluativos referidos a ese foco de atención han sido objeto de reflexión en varias publicaciones en los últimos años:

Tabla 27

Implicaciones del EEES en la evaluación del aprendizaje discente: resumen de aportaciones

REFERENCIAS AL TEMA	EEES: CAMBIOS EN LA EVALUACIÓN DEL APRENDIZAJE DE LAS/OS ESTUDIANTES
McDowell & Sambell (2003)	• Nuevo papel del alumnado en la evaluación. • Cambio en la naturaleza de las tareas evaluativas.
Mateo (2006b, 2008)	• Cambio en el objeto de la evaluación: competencias. • Cambio de la lógica de la evaluación: formativa, centrada en el aprendizaje.
B. Salinas & Cotillas (2007)	• Nuevo papel del alumnado en la evaluación. • Cambio en el objeto de la evaluación: competencias.
Monescillo & Toscano (2007)	• Cambio en el objeto de la evaluación: competencias. • Cambios en los momentos de la evaluación. • Cambio de la lógica de la evaluación (de sancionadora a formativa).
Rué (2007, 2008)	• Nuevo perfil discente y nuevas trayectorias formativas. • Nuevo papel del alumnado en la formación y la evaluación. • Cambio en el objeto de la evaluación: competencias. • Cambios en la distribución de la responsabilidad de la evaluación (de individual a colegiada). • Cambios en las exigencias sociales o acreditativas.
Sabaté et al. (2008)	• Cambios en los agentes participantes de la evaluación. • Cambio en el objeto de la evaluación: competencias.
Font (2010)	• Cambio en el objeto de la evaluación: competencias. • Cambios en los agentes participantes de la evaluación. • Cambios en los momentos de la evaluación.

REFERENCIAS AL TEMA	EEES: CAMBIOS EN LA EVALUACIÓN DEL APRENDIZAJE DE LAS/OS ESTUDIANTES
M. J. García (2010)	• Cambios en la importancia otorgada a los elementos del binomio enseñanza-aprendizaje. • Cambio en el objeto de la evaluación: competencias. • Cambios en la lógica de la evaluación: formativa, centrada en el aprendizaje. • Cambios en las exigencias sociales vinculadas a la evaluación. • Cambios en el uso estratégico de los instrumentos evaluativos. • Cambios en los agentes participantes de la evaluación.
Palomares (2011)	• Cambio en el objeto de la evaluación: competencias. • Cambios en los agentes participantes de la evaluación.
J. Sánchez (2011)	• Cambios en el objeto de la evaluación. • Cambios en los momentos de la evaluación. • Cambios en los usos de la evaluación. • Cambios en los agentes participantes de la evaluación.
V. M. López (2012)	• Cambios en los momentos de la evaluación. • Cambio de la lógica de la evaluación (orientada a la mejora del aprendizaje y la enseñanza). • Cambio en el objeto de la evaluación: competencias.
Guerrero et al., (2013)	• Nuevo papel del alumnado en la evaluación. • Cambios en la lógica de la evaluación: formativa, centrada en el aprendizaje. • Cambio en el objeto de la evaluación: competencias.

La síntesis de las referencias anteriores sugiere que la evaluación del aprendizaje en las universidades europeas está realizando seis ampliaciones o rutas de cambio desde que se iniciara el proceso de convergencia (ver figura 17).

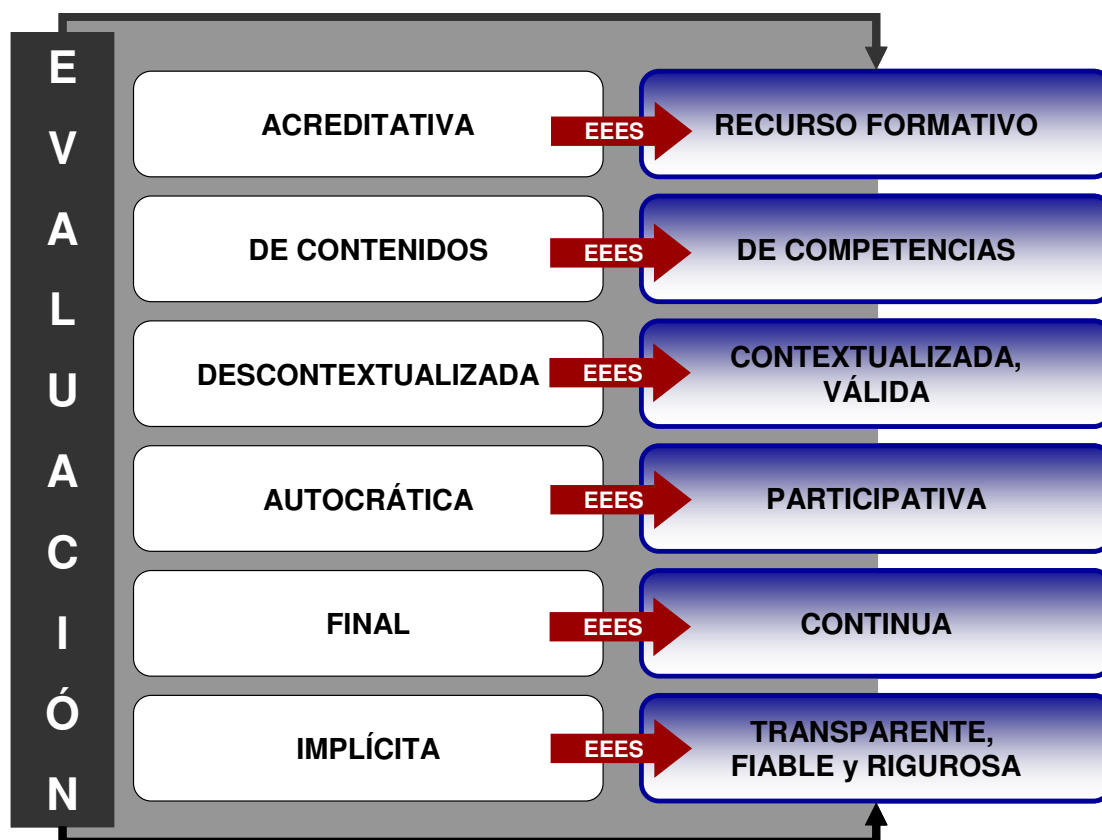


Figura 17. Ampliación de la evaluación del aprendizaje en el contexto formativo del EEES.

Aunque surjan de la mano del proceso de convergencia, las modificaciones planteadas responden a una necesidad más global, apegada a los cambios sociales que están teniendo lugar en las últimas décadas (Rizo, 2004; Rué, 2007). Como indican Palacios y López (2013), plantean exigencias evaluativas muy próximas a las condiciones detalladas desde la perspectiva teórica de la evaluación para el aprendizaje. En consecuencia, la configuración de un nuevo escenario representa una nueva ocasión para modernizar la educación superior en general (Zabalza, 2008), y para introducir mejoras significativas que afecten a los sistemas evaluativos empleados en la misma, en particular.

Avanzar en un nuevo paradigma formativo es incompatible con la persistencia de los modelos de evaluación tradicionales (Alsina, 2011; Cano & Ion, 2012; Villa, 2008). Por ese motivo, las orientaciones que se desprenden de la revisión de esta primera influencia contextual representan una referencia obligada en la construcción de una teoría evaluativa realista y actualizada. La tabla 28 ofrece las principales conclusiones extraídas a tal efecto.

Tabla 28

Exigencias evaluativas del EEES: conclusiones para la definición de una teoría de la evaluación del aprendizaje contextualizada

	ORIENTACIONES Y DIRECTRICES	Notas para la definición de una teoría de la evaluación
EEES	• Nuevas funciones. • Nuevos objetos. • Nuevos agentes. • Nueva cronología. • Mayor exigencia de realismo y validez. • Mayor exigencia de rigor y fiabilidad.	✓ Evaluación continua. ✓ Formativa y sumativa. ✓ Evaluación rigurosa, pública y transparente. ✓ Evaluación participativa. ✓ Evaluación contextualizada (acción y transferencia). ✓ Centrada en el aprendizaje de competencias. ✓ Evaluación globalizada, diversa y divergente. ✓ Evaluación compleja.

1.2.2 Referentes normativos para el desarrollo de la evaluación del aprendizaje en el contexto universitario.

Como ha sido señalado en el capítulo anterior, las regulaciones normativas condicionan poderosamente la evaluación del aprendizaje. Siendo así, la configuración de un modelo de evaluación adecuado al contexto debe partir de su revisión y no desconsiderar su influencia.

Uno de los retos declarados en el paradigma formativo de la convergencia es la promoción de la autonomía académica y pedagógica entre las instituciones formativas de educación superior. En este contexto, las titulaciones de grado o postgrado ofertadas se conciben como proyecto formativo propio de las instituciones que las imparten. Salvadas las condiciones expuestas en el Real Decreto 1393/2007 para la regulación de las enseñanzas universitarias³¹, los centros superiores y universidades tienen potestad para diseñar una oferta formativa acorde a su propia visión estratégica.

Este marco general de autonomía pedagógica también se refleja a la hora de regular la evaluación del aprendizaje discente (Font, 2010). Partiendo de unas orientaciones generales, corresponde a las propias universidades articular reglamentos para regular la praxis evaluativa en su seno. En el contexto formativo seleccionado, dichas orientaciones remiten a un marco normativo conformado por tres referencias complementarias, una de origen europeo y las dos restantes de elaboración estatal:

³¹ Y, en su caso, los requisitos mínimos para las titulaciones vinculadas a profesiones reguladas.

- Los criterios y directrices para la garantía de la calidad aplicables en las universidades que conforman el EEES (ENQA, 2009).
- Las directrices nacionales para el desarrollo de sistemas de garantía interna de la calidad en las universidades. Éstas se presentan en el programa AUDIT elaborado con la participación de la Agencia Nacional de Evaluación de la Calidad y Acreditación, la Àgència per a la Qualitat del Sistema Universitari a Catalunya y la Axencia para a Calidade do Sistema Universitario de Galicia (ANECA, AQU, & ACSUC, 2007a, 2007b).
- El Estatuto del estudiante universitario, que regula derechos y deberes del alumnado universitario español (Real Decreto 1791/2010) y que incorpora regulaciones sobre su participación en la evaluación.

Como detallan los próximos apartados, su consideración conjunta conlleva una serie de implicaciones evaluativas de las que da cuenta la tabla 29. La delimitación del modelo ITC las incorpora en su desarrollo para favorecer un buen ajuste a las condiciones del contexto normativo de referencia:

Tabla 29

Referentes normativos para el ejercicio de la evaluación: conclusiones para la definición de una teoría de la evaluación del aprendizaje contextualizada

	ORIENTACIONES Y DIRECTRICES	Notas para la definición de una teoría de la evaluación
ENQA	• Cualificación de los agentes evaluadores (ejercicio profesional). • Adecuación pedagógica de los instrumentos de evaluación. • Publicidad y transparencia de criterios, normas y condiciones. • Cumplimiento coherente de las normas y compromisos.	✓ <i>Instrumentos adecuados desde la perspectiva pedagógica.</i> ✓ <i>Información pública y transparente.</i> ✓ <i>Desarrollo o puesta en práctica coherente.</i>
ANECA	• Regulación de la evaluación del aprendizaje que garantice su claridad y transparencia. • Empleo de los resultados de aprendizaje como evidencia a nivel institucional.	✓ <i>Evaluación consciente (cualificada, profesional).</i> ✓ <i>Evaluación objetiva.</i> ✓ <i>Evaluación continua.</i>
Real Decreto 1791/2010	• Publicidad y transparencia de criterios, normas y condiciones. • Puesta en práctica coherente. • Evaluación objetiva, continua y formativa.	✓ <i>Sumativa y formativa.</i>

Este primer recorrido por los principales referentes normativos sugiere una presencia limitada de la evaluación del aprendizaje. Las alusiones constatadas son escasas y hacen

referencia a aspectos muy concretos de la práctica evaluativa. El uso pedagógico de la evaluación encuentra un reflejo limitado en las tres referencias; en su lugar, se insiste especialmente en los procedimientos de evaluación aplicados y en el establecimiento público de criterios y normas claras de actuación. Estas conclusiones coinciden con las publicadas por Ibarra y Rodríguez (2010), Padilla et al. (2010) y Porto (2010), que también evidencian la escasa regulación de la evaluación del aprendizaje, más allá del calendario, de las condiciones de examen o de alusiones a las tasas de éxito, abandono, graduación o duración media de los estudios.

1.2.2.1 Criterios y directrices europeas para la garantía de la calidad.

Con el objetivo de asentar unas bases comunes entre las universidades de los países miembro, la ENQA elabora una serie de *criterios y directrices para la garantía de la calidad*. En su tercera edición, dichos criterios aparecen divididos en torno a tres espacios (ENQA, 2009): la garantía de la calidad interna, la garantía de la calidad externa y las agencias para la garantía externa de la calidad. En cada uno de ellos, se realiza una breve descripción de los criterios³² incluidos y se presentan una serie de directrices para orientar su consecución.

La evaluación del aprendizaje aparece recogida en el criterio 1.3, dentro del primer apartado. Este establece que el alumnado debe ser evaluado utilizando criterios, normativas y procedimientos públicos y aplicados de forma coherente (ENQA, 2009). Reconocido el importante papel que juega en la calidad de la educación superior y en la evolución del alumnado, insta a las universidades a realizar un uso profesional de la evaluación, basado en el amplio conocimiento disponible sobre los procesos de prueba y examen. También la presenta como una fuente destacada de información institucional sobre la efectividad docente.

Para favorecer su consecución, el criterio se acompaña de nueve directrices de actuación que, en conjunto, desarrollan cuatro orientaciones a incorporar (ver tabla 29):

- *Adecuación pedagógica.* Deben disponerse instrumentos capaces de medir los aprendizajes esperados, desarrollados y aplicados con fines pedagógicos o acreditativos.
- *Cualificación de los agentes evaluadores.* Las/os evaluadoras/es deben ser conscientes de la importancia de la función evaluativa y de las consecuencias de la

³² Estándares en la versión original.

normativa que regula la realización de exámenes. Una de las directrices explicita que, siempre que sea posible, la evaluación no debería implicar a un único agente evaluador.

- *Claridad.* La evaluación del aprendizaje debe regirse por normas claras que incluyan las condiciones generales de participación, así como la regulación de las situaciones de excepcionalidad (ausencias por enfermedad u otras circunstancias atenuantes).
- *Cumplimiento coherente.* La evaluación debe desarrollarse en coherencia con la planificación prevista y la normativa. También establece la realización de inspecciones administrativas que supervisen que así suceda.

1.2.2.2 Orientaciones nacionales para la garantía de la calidad interna.

Las exigencias de transparencia del EEES de un lado, y la ordenación de los procesos de evaluación y certificación de la enseñanza del otro, sitúan a las universidades españolas ante el reto de elaborar sus propios sistemas de garantía interna de la calidad. El Programa AUDIT responde a esta necesidad y surge para brindar apoyo en la tarea. En concreto, establece una serie de criterios y directrices para la garantía de la calidad, de modo que puedan ser adoptados como marco común de referencia en todas las universidades y agencias de calidad del estado (ANECA, AQU, & ACSUC, 2007a). Dichos criterios, surgen de la consideración de las directrices europeas (ENQA, 2009), así como de las ofrecidas por el Ministerio de Educación y Ciencia para la elaboración de los nuevos títulos de grado y postgrado (ANECA, AQU, & ACSUC, 2007b).

En suma, el programa AUDIT establece siete directrices (una general, y seis específicas) acompañadas de descriptores para orientar su consecución. La revisión de las mismas nos lleva a señalar referencias muy difusas en torno a la evaluación del aprendizaje. Detallados en la tabla 30, los descriptores que más directamente aluden a ella, pueden ordenarse en dos bloques: aquellos orientados a la información y transparencia (B) y los que presentan la evaluación del aprendizaje como una fuente de información relevante sobre el desarrollo y la efectividad profesional de la institución (A).

Tabla 30

Directrices para el desarrollo de sistemas de garantía interna de la calidad del programa AUDIT.
 Descriptores referidos a la evaluación del aprendizaje (ANECA, AQU, & ACSUC, 2007b)

CÓD.	Descriptores referidos a la evaluación del aprendizaje	BLOQUE
1.2.1	Disponer de sistemas de información que le permitan conocer y valorar las necesidades del Centro en materia de: enseñanza y evaluación de los aprendizajes (entre otras).	A
1.5.2	Contar con sistemas de recogida de información que faciliten datos relativos a los resultados del aprendizaje , de la inserción laboral y de la satisfacción de los grupos de interés.	
1.2.3	Establecer mecanismos que regulen las directrices que afectan a las/os estudiantes: reglamentos (exámenes , sanciones, petición de certificaciones, convalidaciones, etc.), normas de uso (de instalaciones), calendarios, horarios y beneficios que ofrece la Universidad.	B
1.2.7	Indicar el procedimiento (cómo, quién, cuándo) seguido para rendir cuentas sobre los resultados del aprendizaje de las/os estudiantes	
1.6.2	Determinar cuál es el procedimiento establecido para informar a los grupos de interés (incluyendo los distintos niveles de la estructura organizativa del Centro) acerca de: metodologías de enseñanza, aprendizaje y evaluación (incluidas las prácticas externas (entre otras).	

Nota: La primera columna de la tabla (CÓD.) recoge la codificación original de descriptores en el programa. Las referencias a la evaluación aparecen destacadas en negrita.

La revisión de estos descriptores sugiere dos conclusiones relevantes con respecto a la evaluación del aprendizaje, ambas contempladas en la tabla 29.

- La importancia de garantizar la publicidad y transparencia del proceso evaluativo.
- El empleo de los resultados de la evaluación del aprendizaje como evidencias de avance y logro institucional.

La imagen que devuelve de la evaluación el programa AUDIT parece más dirigida a la función acreditativa que a su misión pedagógica. Centrada en la rendición de cuentas y en la regulación de los derechos y deberes básicos de agentes evaluados y evaluadores, en los descriptores no hay ninguna referencia al valor pedagógico de la evaluación. Esta constatación no es consistente con las sugerencias de Stiggins y Duke (2008), Banta et al. (1996) o Ion y Cano (2011), que sostienen que la mejora de la evaluación del aprendizaje precisa de liderazgo institucional expreso. Tampoco con los testimonios que sitúan la evaluación del aprendizaje como factor clave de la calidad universitaria (Gibbs, 2010; Kaneko, 2008; Villa, 2004).

1.2.2.3 Estatuto del estudiante universitario.

El Real Decreto 1791/2010 presenta el estatuto que rige los principales derechos y deberes del alumnado en su participación en la comunidad universitaria. A lo largo del documento hay varias referencias a la evaluación del aprendizaje discente, pero es de particular interés la aportación del artículo 25, dedicado íntegramente a su desarrollo:

Artículo 25. Evaluación de los aprendizajes del estudiante.

1. La evaluación del rendimiento académico de los estudiantes responderá a criterios públicos y objetivos y tenderá hacia la evaluación continua, entendida como herramienta de corresponsabilidad educativa y como un elemento del proceso de enseñanza-aprendizaje que informa al estudiante sobre su proceso de aprendizaje.
2. La evaluación se ajustará a lo establecido en los planes docentes de las materias y asignaturas aprobados por los departamentos.
3. Los calendarios de fechas, horas y lugares de realización de las pruebas, incluidas las orales, serán acordados por el órgano que proceda, garantizando la participación de los estudiantes, y atendiendo a la condición de que éstos lo sean a tiempo completo o a tiempo parcial.
4. La programación de pruebas de evaluación no podrá alterarse, salvo en aquellas situaciones en las que, por imposibilidad sobrevenida, resulte irrealizable según lo establecido. Ante estas situaciones excepcionales, los responsables de las titulaciones realizarán las consultas oportunas, con el profesorado y los estudiantes afectados para proceder a proponer una nueva programación de acuerdo con lo previsto en la normativa autonómica y de las propias universidades.
5. Los estudiantes que, por motivos de asistencia a reuniones de los órganos colegiados de representación universitaria, o por otros motivos previstos en sus respectivas normativas, no puedan concurrir a las pruebas de evaluación programadas, tendrán derecho a que les fije un día y hora diferentes para su realización. Las Universidades velarán, conforme a su normativa y a la autonómica, por no hacer coincidir las reuniones con los periodos de exámenes ni con los días de estudio previos.
6. En la programación de los sistemas de evaluación se evitará, de conformidad con lo establecido en la normativa autonómica y de la propia universidad, que un estudiante sea convocado a pruebas de carácter global de distintas asignaturas del mismo curso en un plazo inferior a veinticuatro horas. En todo caso y de acuerdo con la anterior normativa, tendrá derecho a que la realización de las pruebas de carácter global correspondientes no le coincida en fecha y hora. En el caso de las universidades a distancia, esta programación se ajustará a su metodología docente y de evaluación.
7. En cualquier momento de las pruebas de evaluación, el profesor podrá requerir la identificación de los estudiantes asistentes, que deberán acreditarla mediante la exhibición de su carné de estudiante, documento nacional de identidad, pasaporte o, en su defecto, acreditación suficiente a juicio del evaluador.
8. Los estudiantes tendrán derecho a que se les entregue a la finalización de las pruebas de evaluación un justificante documental de haberlas realizado.

La imagen que devuelve de la evaluación este estatuto es consistente con la tendencia advertida en las dos referencias normativas anteriores (tabla 29). La mayor parte de los puntos hacen referencia a la información pública sobre los criterios y condiciones de la evaluación y al desarrollo coherente de las mismas en la práctica. Se orienta prioritariamente a la función social de la evaluación, regulando los derechos del alumnado

en la realización de pruebas específicas y las vicisitudes que puedan tener lugar en su desarrollo. El único punto que aporta un elemento distintivo a este discurso es el primero, donde se expresa el deseo de consolidar una práctica evaluativa educativa, objetiva y continua.

1.2.3 La evaluación del aprendizaje en la formación básica de maestras/os.

La evaluación del aprendizaje resulta especialmente relevante en el marco de las titulaciones asociadas a la formación básica de maestras/os de Educación Infantil y Primaria. Dos son los motivos que avalan tal afirmación:

- Por una parte, la de maestra/o es una profesión regulada, con implicaciones formativas ordenadas legalmente. Ello conlleva una responsabilidad evaluativa añadida que debe considerarse en su ejercicio.
- Por otra parte, en las titulaciones acotadas existe un compromiso explícito con la capacitación evaluativa de las/os estudiantes, futuras/os maestras/os y evaluadoras/es. En consecuencia, además de representar un componente clave del proceso de enseñanza-aprendizaje, la evaluación del aprendizaje figura entre los resultados de aprendizaje profesional a lograr por el alumnado.

Desarrolladas en los siguientes apartados, la consideración de ambas circunstancias da lugar a la incorporación de unos rasgos evaluativos precisos en el modelo ITC. Los anticipa la tabla 31:

Tabla 31

Implicaciones evaluativas en la formación básica de maestras/os: conclusiones para la definición de una teoría de la evaluación del aprendizaje contextualizada

	ORIENTACIONES Y DIRECTRICES	CONCLUSIONES (TEORÍA DE LA EVALUACIÓN)
PROFESIÓN REGULADA	• Acreditación de logros transparente y rigurosa para los procesos de reconocimiento de títulos y movilidad. • Promoción estratégica de la capacitación evaluativa dispuesta en las directrices por medio del modelado.	✓ <i>Evaluación coherente (modelado pedagógico).</i> ✓ <i>Evaluación rigurosa y transparente.</i>
EVALUACIÓN COMO OBJETIVO	• Promoción estratégica de la capacitación evaluativa por medio del modelado.	

1.2.3.1 *La formación de maestras/os de Educación Infantil y Primaria en el contexto de las profesiones reguladas.*

En virtud de una mayor autonomía institucional y pedagógica, las universidades cuentan con la posibilidad de ofertar sus propias titulaciones sin atenerse a los límites de un catálogo predeterminado de titulaciones. La única condición es que se cumplan los requisitos generales para la verificación del título por parte de la ANECA, lo que autorizaría a su impartición.

En este contexto, cobra especial relevancia el concepto de profesión regulada, relacionado con la libre circulación de personas y servicios dentro de los estados miembros que conforman el EEES. El Real Decreto 1837/2008 regula su definición, integrando las directrices europeas:

A los exclusivos efectos de la aplicación del sistema de reconocimiento de cualificaciones regulado en este real decreto, se entenderá por `profesión regulada´ la actividad o conjunto de actividades profesionales para cuyo acceso, ejercicio o modalidad de ejercicio se exija, de manera directa o indirecta, estar en posesión de determinadas cualificaciones profesionales, en virtud de disposiciones legales, reglamentarias o administrativas. A estos efectos, las profesiones y las actividades que entran dentro del ámbito de aplicación del sistema de reconocimiento de cualificaciones según la definición anterior son las que se relacionan en el anexo VIII sin que de dicha inclusión puedan derivarse otros efectos fuera de ese ámbito. (Art. 4.1).

Las profesiones de maestra/o de Educación Infantil y Educación Primaria están presentes en el citado anexo y, en consecuencia, se incluyen dentro del bloque de las reguladas. Este reconocimiento introduce una serie de condiciones en cuanto a su formación básica, repercutiendo en el diseño e implantación de las titulaciones. Así lo explicita el Real Decreto 1393/2007 sobre la ordenación de las enseñanzas universitarias oficiales:

Cuando se trate de títulos que habiliten para el ejercicio de actividades profesionales reguladas en España, el Gobierno establecerá las condiciones a las que deberán adecuarse los correspondientes planes de estudios, que además deberán ajustarse,

en su caso, a la normativa europea aplicable. Estos planes de estudios deberán, en todo caso, diseñarse de forma que permitan obtener las competencias necesarias para ejercer esa profesión. A tales efectos la Universidad justificará la adecuación del plan de estudios a dichas condiciones. (Art. 12.9).

Las condiciones a las que hace referencia el artículo anterior se concretan en el cumplimiento de unas directrices generales sobre la denominación, la estructuración y el contenido de los planes de estudio. Presentadas en el siguiente apartado, las Órdenes ECI/3854/2007 y ECI/3857/2007 establecen tales directrices para los títulos de Maestra/o de Infantil y Primaria, respectivamente.

El carácter normativo de estas regulaciones, no sólo debería reflejarse en la estructura interna de las titulaciones y en el establecimiento de resultados de aprendizaje, sino también en su evaluación. Estas implicaciones inciden tanto en la responsabilidad pedagógica de la evaluación como en la técnica y así consta en el resumen efectuado en la tabla 29:

- *Responsabilidad técnica de rigor y transparencia.* Para garantizar el cumplimiento de los mínimos comunes que posibilitan la movilidad de las profesiones reguladas entre los estados miembros, es indispensable que se vele por el cumplimiento fiel de las directrices establecidas. Además del desarrollo de prácticas evaluativas alineadas desde la perspectiva más pedagógica, ello exige garantizar el rigor y transparencia de la evaluación del aprendizaje acreditado en el alumnado.
- *Responsabilidad pedagógica de la coherencia (modelado).* La propia vivencia evaluativa experimentada durante la formación básica puede entorpecer o reforzar el desarrollo de las competencias previstas en las/os futuras/os maestras/os, en especial, aquellas asociadas a su futura experiencia como evaluadoras/es (V. M. López, 2004; Palacios & López, 2013). Ello exige estimular deliberadamente prácticas acordes a lo dispuesto en las directrices por medio de la evaluación.

1.2.3.2 La evaluación del aprendizaje como competencia docente destacada y resultado de aprendizaje.

Como ya ha sido expuesto a lo largo del capítulo tercero, toda/o estudiante desarrolla aprendizajes implícitos sobre la evaluación a lo largo de su trayectoria formativa. En las titulaciones acotadas, el alumnado debe construir además un aprendizaje explícito

que le permita desarrollar su competencia evaluativa para el desempeño profesional docente. Así lo reflejan los requisitos de verificación de los títulos universitarios para el ejercicio de la profesión de Maestra/o en Educación Infantil y Primaria (Orden ECI/3854/2007 y Orden ECI/3857/2007).

Las directrices normativas establecen para cada profesión un perfil de 12 competencias específicas y la evaluación del aprendizaje aparece igualmente representada en ambos casos. Según recoge la tabla 32, en cada titulación se detectan tres alusiones directas a la capacitación evaluativa, que se desarrollan en las competencias atribuidas al módulo de formación básica y al didáctico-disciplinar.

Tabla 32

La competencia evaluativa en los decretos reguladores para la formación básica de maestras/os de Educación Infantil y Primaria

	EDUCACIÓN PRIMARIA Orden ECI/3857/2007	EDUCACIÓN INFANTIL Orden ECI/3854/2007
COMPETENCIAS (n=12)	1. Conocer las áreas curriculares de la Educación Primaria, la relación interdisciplinar entre ellas, los criterios de evaluación y el cuerpo de conocimientos didácticos en torno a los procedimientos de enseñanza y aprendizaje respectivos. 2. Diseñar, planificar y evaluar procesos de enseñanza y aprendizaje, tanto individualmente como en colaboración con otros docentes y profesionales del centro. 5. Fomentar la convivencia en el aula y fuera de ella, resolver problemas de disciplina y contribuir a la resolución pacífica de conflictos. Estimular y valorar el esfuerzo, la constancia y la disciplina personal en las/os estudiantes.	1. Conocer los objetivos, contenidos curriculares y criterios de evaluación de la Educación Infantil. 4. Fomentar la convivencia en el aula y fuera de ella y abordar la resolución pacífica de conflictos. Saber observar sistemáticamente contextos de aprendizaje y convivencia y saber reflexionar sobre ellos. 6. Conocer la evolución del lenguaje en la primera infancia, saber identificar posibles disfunciones y velar por su correcta evolución.
FORMACIÓN BÁSICA	MATERIA 02. Diseñar, planificar y evaluar la actividad docente y el aprendizaje en el aula. Conocer y aplicar metodologías y técnicas básicas de investigación educativa y ser capaz de diseñar proyectos de innovación identificando indicadores de evaluación.	MATERIA 02. Identificar dificultades de aprendizaje, disfunciones cognitivas y las relacionadas con la atención... MATERIA 06. Comprender que la observación sistemática es un instrumento básico para poder reflexionar sobre la práctica y la realidad, así como contribuir a la innovación y a la mejora en Educación Infantil. Dominar las técnicas de observación y registro. Abordar análisis de campo mediante metodología observacional utilizando tecnologías de la información, documentación y audiovisuales. Saber analizar los datos obtenidos, comprender críticamente la realidad y elaborar un informe de conclusiones.

DIDÁCTICO DISCIPLINAR	<u>MATERIAS 04-09.</u> Desarrollar y <i>evaluar</i> contenidos del currículo mediante recursos didácticos apropiados y promover la adquisición de competencias básicas en las/os estudiantes.	<u>MATERIAS 08-10.</u> Ninguna explícita: • <i>Elaboración de propuestas didácticas</i> • <i>Conocimiento del currículo</i> • <i>Valoración de...</i>
PRACTICUM	X	X

Nota: Las referencias a la evaluación del aprendizaje aparecen destacadas en negrita.

Aunque la representación cuantitativa de la evaluación sea igual en ambas titulaciones, su desarrollo es diferente en cada caso. No es objeto de esta investigación valorar la formación básica que reciben las/os maestras/os de Educación Infantil y Primaria para el futuro cumplimiento de sus funciones evaluativas, aunque este análisis vertical y superfluo suscita ciertos interrogantes de partida:

- La presencia de la evaluación del aprendizaje en las directrices para la preparación de maestras/os de Educación Infantil es difusa. Tanto en la formulación de competencias como en los tres tipos de módulo delimitados para su desarrollo, se hace referencia a la observación, a la identificación o al diagnóstico. La asimilación implícita y generalizada que se produce entre evaluación y calificación puede ser una de las razones de esta evitación, lo que explicaría la búsqueda de términos que enfatizen más el carácter formativo que corresponde a la etapa.
- En Educación Primaria la evaluación figura como un componente más del currículo y parece estrechamente vinculado a las áreas, reforzando la tradicional visión disciplinar. Esta constatación se hace especialmente evidente en el módulo didáctico- disciplinar, donde la evaluación refiere más a los contenidos que a las competencias a desarrollar.
- En ambas titulaciones, resulta llamativo que la descripción de competencias referidas al Practicum no incorpore ninguna referencia expresa a la función evaluativa.

A nivel internacional, son varias las voces que alertan sobre una preparación insuficiente e inadecuada de las/os maestras/os para el ejercicio de la evaluación (Gullickson, 1993/2012; JCSEE, 2007; Mertler, 2009; Sanders & Vogel, 2012/1993; Stiggins, 2012/1993, 1998; Stiggins & Conklin, 1988; Wise & Lukin, 2012/1993) y en el ámbito español parece pertinente sugerir una línea de investigación para indagar a este respecto. Más aun considerando el testimonio de Stiggins (2002), quien sostiene que la escasa

representación formativa de esta competencia explica que el profesorado no esté capacitado para emplear la evaluación como recurso para el aprendizaje.

Con independencia de la cantidad y calidad de la oferta formativa para la capacitación evaluativa, su presencia entre las competencias a desarrollar implica nuevamente la necesidad de coherencia pedagógica por parte del profesorado (tabla 29). De ahí que Stiggins (1998) destaque el modelado docente de prácticas evaluativas deseables como uno de los elementos clave en la formación básica de maestras/os³³.

1.3 Antecedentes en el establecimiento de criterios, principios o estándares para la evaluación del aprendizaje universitario.

Junto al enfoque teórico de referencia y las particularidades del contexto formativo al que se orienta la investigación, la consulta de los antecedentes relevantes en el tema representa la tercera referencia para la configuración de nuestra particular teoría de la evaluación (figura 15).

La literatura específica da cuenta de numerosos precedentes que han tratado de sistematizar los rasgos a los que debería ajustarse la práctica evaluativa. Para acotar la consulta a las propuestas que puedan ofrecer una orientación más válida y certera, hemos centrado la atención en aquellas específicamente referidas a la evaluación del aprendizaje universitario. Del mismo modo, se han primado las referencias con mayor impacto en la literatura específica, procedentes de autoras/es o asociaciones de reconocido prestigio.

Las aportaciones seleccionadas han sido agrupadas en tres bloques en función de su grado de sistematización o desarrollo (ver tabla 33). En el primer bloque se muestran decálogos o manifiestos que han tratado de dibujar condiciones para la buena práctica evaluativa en la universidad desde una visión amplia y fundamentalmente teórica. El segundo bloque integra propuestas estructuradas en las que se delimitan criterios más concretos para conducir la evaluación del aprendizaje. Por último, la mayor sistematización se produce en el tercer bloque, donde se incluyen dos propuestas estandarizadas.

³³ Sobre esta referencia nos detendremos en el próximo capítulo, al analizar los antecedentes en la delimitación de competencia evaluativa docente.

Tabla 33

Antecedentes en la sistematización de un modelo o teoría de la evaluación del aprendizaje

ANTECEDENTES: PROPUESTAS PREVIAS CONSIDERADAS	
Decálogos y manifiestos	• <i>Manifiesto sobre evaluación</i> (Brown et al., 1996). • <i>14 normas para una mejor evaluación en la educación superior</i> (Ramsden, 2003). • <i>Condiciones para que la evaluación promueva el aprendizaje</i> (Gibbs & Simpson, 2009). • <i>REAP Project (Re-Engineering Assessment Practices in Scottish Higher Education)</i> (2005-2007). • <i>Principios para la buena evaluación y retroalimentación</i> (Nicol 2007, 2009). • <i>Estándares de evaluación: Manifiesto para el cambio</i> (Weston Manor Group- WMG- Assessment Standards Knowledge Exchange- ASKe, 2007). • <i>Assessment 2020</i> (Boud & asociados, 2010). • <i>Observaciones sobre feedback evaluación</i> (National Union of Students- NUS, 2010). • <i>Guía para la evaluación para el aprendizaje</i> (2ª edición) (Center for Learning Excellence, University of Bedfordshire- CLE-UB, 2011). • <i>Criterios para prácticas evaluativas de calidad</i> (Grup d'Avaluació de la Pràctica Acadèmica- Recerca i Innovació en Metodologies de l'Aprenentatge de la Universitat Politècnica de Catalunya RIMA- Grupo GRAPA-RIMA, 2012).
Propuestas estructuradas	• <i>Principios para la buena práctica de la evaluación de las/os estudiantes</i> (Astin et al., 1992 - AAHE-American Association of Higher Education). • <i>Indicadores de la evaluación efectiva para las universidades Australianas</i> (CSHE- Centre for the Study of Higher Education, 2002). • <i>Código de actuación para el aseguramiento de la calidad académica y los estándares en la educación superior</i> (Quality Assurance Agency for Higher Education- QAA, 2011).
Propuestas estandarizadas	• <i>Estándares para la evaluación de los aprendizajes de las/os estudiantes</i> (The Joint Committee on Standards for Educational Evaluation, 2007). • <i>Requisitos para la evaluación del aprendizaje del alumnado</i> (National Center for Research on Evaluation, Standards & Student Testing- CRESST).

La consulta de esta selección de referencias previas contribuye a definir una propuesta sólida y bien fundamentada para la evaluación del aprendizaje universitario. La tabla 34 anticipa las aportaciones que ha dejado para la definición del modelo ITC su revisión, ordenadas nuevamente por bloques; las contribuciones extraídas de los decálogos y propuestas estructuradas son directas, mientras que los estándares proporcionan un marco general de actuación sobre el que desarrollar la evaluación del aprendizaje.

Tabla 34

Antecedentes en la sistematización de rasgos evaluativos deseables: conclusiones para la definición de una teoría de la evaluación del aprendizaje fundamentada

	ORIENTACIONES Y DIRECTRICES	CONCLUSIONES (TEORÍA DE LA EVALUACIÓN)
DECÁLOGOS Y MANIFIESTOS	- Planificación estratégica de la evaluación. - Publicidad y transparencia. - Comunicación efectiva de resultados. - Orientación a la mejora. - Implicación del alumnado. - Presencia de propiedades relevantes. - Creación de comunidades de aprendizaje.	✓ Evaluación alineada. ✓ Evaluación orientada al aprendizaje. ✓ Evaluación continua y pedagógica. ✓ Evaluación sensible con el proceso de aprendizaje. ✓ Adecuación y variedad de instrumentos y tareas. ✓ Evaluación compleja y contextualizada. ✓ Evaluación pública y transparente. ✓ Interiorización de expectativas altas y claras de aprendizaje. ✓ Feedback de calidad. ✓ Regulación del aprendizaje.
PROPUESTAS ESTRUCTURADAS	- Planificación estratégica de la evaluación. - Publicidad y transparencia. - Comunicación efectiva de resultados. - Implicación del alumnado. - Política institucional evaluativa. - Evaluación profesional. - Precisión de la evaluación. - Puesta en práctica coherente. - Equidad.	✓ Regulación de la enseñanza. ✓ Metaevaluación. ✓ Evaluación participativa. ✓ Evaluación fiable y rigurosa. ✓ Evaluación válida. ✓ Evaluación equitativa. ✓ Evaluación coherente con normas, estándares y directrices institucionales. ✓ Evaluación generadora de comunidades de aprendizaje. ✓ Implementación coherente. ✓ Evaluación profesional (consciente, reflexiva).

Adecuación, precisión, viabilidad y utilidad (JCSEE, 2007)

Validez, credibilidad, justicia y utilidad (CRESST)

Los tres siguientes apartados detallan las conclusiones derivadas de cada bloque de antecedentes.

1.3.1 Decálogos y manifiestos para la evaluación del aprendizaje universitario

Este primer bloque incluye los antecedentes más amplios y teóricos respecto al tema. En forma de decálogo o manifiesto, los trabajos que lo integran proporcionan referencias generales para definir cómo debe tener lugar la evaluación del aprendizaje en la universidad.

En total han sido seleccionadas 10 propuestas, introducidas brevemente a continuación en orden cronológico:

1. *Manifiesto sobre evaluación* (Brown et al., 1996). Como se desprende de su propio nombre, esta primera referencia representa un manifiesto, a saber, una declaración pública de propósitos o doctrina (Real Academia Española, 2014). De forma clara y concisa, la propuesta concentra 10 puntos clave para encauzar la mejora de la evaluación del aprendizaje universitario. Formulada hace casi dos décadas, la propuesta sigue teniendo vigencia en el contexto universitario actual y se muestra próxima a otras de las referencias más recientes.
2. *14 normas para una mejor evaluación en la educación superior* (Ramsden, 2003). Prestigioso profesor e investigador de la Universidad de Sidney, Paul Ramsden (2003) considera la evaluación como un instrumento de comprensión docente, que juega un papel crucial en la determinación de la calidad del aprendizaje discente. Para comprobarlo, realizó un estudio de casos en los que exploraba el empleo efectivo de la evaluación por parte de profesorado de titulaciones distintas. Este conjunto de normas representa la conclusión de tal experiencia. Directamente orientadas al profesorado, ofrecen pautas para guiar la praxis evaluativa y mejorar con ello el aprendizaje del alumnado.
3. *Condiciones para que la evaluación promueva el aprendizaje* (Gibbs & Simpson, 2009). Ampliamente referida en la literatura, esta propuesta sintetiza 10 claves fundamentales para promover el aprendizaje del alumnado universitario por medio de la evaluación. En su desarrollo concentra los puntos centrales del enfoque teórico de la evaluación para el aprendizaje, aquel que se asume como referencia en la presente investigación. Su publicación original se produjo en 2004; en una revisión posterior, Gibbs (2006) añadió un nuevo punto y organizó las once condiciones resultantes en cinco núcleos complementarios.
4. *REAP Project* (2005-2007). Se trata de un proyecto desarrollado entre las universidades escocesas de Strathclyde, Glasgow y Glasgow Caledonian bajo la dirección del profesor David Nicol. Su principal propósito es contribuir a la autorregulación del alumnado universitario, considerada una capacidad clave para posibilitar el aprendizaje a lo largo de la vida. Para potenciar el control, la gestión y la autodirección del aprendizaje entre las/os estudiantes, adopta el feedback y la evaluación formativa como principales ejes de intervención.

El proyecto se basa en 11 principios evaluativos compartidos en las tres instituciones implicadas y ampliados en cada caso particular³⁴. Cuatro de ellos hacen referencia al establecimiento de tareas o contratos de aprendizaje, y derivan directamente de la propuesta de condiciones de Gibbs y Simpson (2009). Los siete principios restantes representan una reelaboración de las orientaciones de Nicol y Macfarlane (2004, 2006) para el ejercicio del feedback³⁵.

De modo similar, el proyecto TESTA (Transforming the Experience of Students Through Assessment) trata de mejorar la calidad del aprendizaje por medio de la evaluación en el contexto del Reino Unido (Jessop, 2012). Fundado por la Higher Education Academy (HEA) en 2009 y liderado desde la Universidad Winchester, dicho proyecto asume como marco de referencia las mismas condiciones evaluativas determinadas por Gibbs y Simpson (2009), por eso no ha sido incluido en el listado de referentes.

5. *Principios para la buena evaluación y retroalimentación* (Nicol 2007, 2009). Coherente con las orientaciones del enfoque teórico de la evaluación para el aprendizaje, esta propuesta destaca por relevante e influyente. Así por ejemplo, ha condicionado de forma directa otras contribuciones destacables como la declaración de principios del *Joint Information Systems Committee* (JISC, 2010) sobre la evaluación efectiva en la era digital.

La primera propuesta de Nicol (2007) enumeraba 10 condiciones para la buena evaluación y retroalimentación, partiendo de siete principios formulados en un trabajo previo (Nicol & Macfarlane, 2004, 2006). Dos años más tarde, el autor sometió a revisión la propuesta, a la luz de la experiencia acumulada en la dirección del Proyecto REAP y de su participación en la determinación de criterios de calidad para la educación superior del Reino Unido³⁶. En esta última edición, se suman dos nuevos principios completando la docena.

6. *Estándares de evaluación: Manifiesto para el cambio* (WMG-ASKe, 2007). Es una propuesta clara y manejable para la mejora de las deficiencias asociadas a la evaluación del aprendizaje universitario en el Reino Unido. Su definición es la

³⁴ A modo de ejemplo, en <http://www.reap.ac.uk/reap/resourcesPrinciples.html> está disponible la reelaboración practicada en la Universidad de Strathclyde de Glasgow a partir de los 11 principios de referencia.

³⁵ Estas condiciones representan una primera aproximación a las condiciones presentadas por Nicol (2007, 2009) y expuestas a continuación.

³⁶ Dicha contribución se origina desde la QAA (*Quality Assurance Agency*), cuyas contribuciones para la orientación de las prácticas evaluativas se exponen en siguiente bloque.

respuesta a la situación puesta de manifiesto en el informe Burgess (Burgess, 2007). Desde su publicación, es adoptada como referencia práctica por varias universidades dentro y fuera del Reino Unido³⁷.

El ASKe fue creado en el año 2005 como un centro de excelencia en enseñanza y aprendizaje con sede en la Oxford Brookes University Business School. Para la formulación de esta propuesta estableció un foro internacional conformado por 37 expertos en evaluación universitaria, el Grupo Weston Manor. El manifiesto resultante recoge de forma sintética seis líneas para promover el cambio decidido de la evaluación, todas ellas compatibles con las orientaciones que se desprenden del enfoque teórico de la evaluación para el aprendizaje.

7. *Assessment 2020* (Boud & asociados, 2010). Se trata de un manifiesto que concentra siete propuestas concretas para orientar la mejora de la evaluación del aprendizaje en la educación superior. De reconocido prestigio internacional, sugiere un planteamiento evaluativo apegado a la conceptualización de la evaluación para el aprendizaje. La propuesta surge en Australia, donde obtiene amplio respaldo y seguimiento³⁸. A pesar de la distancia, los principios en los que se inspira son compatibles con el paradigma formativo del EEES y, en consecuencia, aplicables al contexto formativo europeo.
8. *Observaciones sobre feedback y evaluación* (NUS, 2010). Esta sexta referencia presenta un decálogo con orientaciones para mejorar la praxis evaluativa en Reino Unido. De todas las referencias consideradas es la única que contiene expresamente el punto de vista discente, ya que la firma la Unión Nacional de Estudiantes (NUS, 2010). Su propuesta responde a las deficiencias evaluativas puestas de manifiesto en 2009 en el National Student Survey (NSS), de modo que parte de una base sólida y fundamentada. Las propuestas que incorpora están muy centradas en el feedback, principal inquietud para el alumnado británico en dicha edición del estudio, así como en la siguiente (Buckley, 2012).
9. *Guía para la evaluación para el aprendizaje* (2ª edición) (CLE-UB, 2011). Partiendo del Manifiesto para el cambio del ASKe (WMG, 2007) y de las directrices generales

³⁷ Fuera de su contexto inmediato hemos hallado referencias al mismo en la Universidad de Canberra o en el Assessment Resource at Hong Kong University (AR@HKU), por ejemplo.

³⁸ Las Universidades de Canberra, Victoria o la Universidad Tecnológica de Sydney, por ejemplo, exponen directamente el documento al presentar su política evaluativa.

de la QAA³⁹, el Centro para la Excelencia del Aprendizaje de la Universidad de Bedfordshire desarrolla una completa guía para orientar y revisar la praxis evaluativa que se desarrolla en su seno. La primera edición de la misma se realizó en 2009 contando con el asesoramiento de los expertos del ASKe. Incorpora quince requisitos a los que debe ajustarse la evaluación para promover el aprendizaje, organizados en cinco bloques. A pesar de representar una propuesta institucional particular, esta guía ha sido seleccionada como referente por incorporar, en una estructura clara y ordenada, orientaciones procedentes de dos antecedentes distintos considerados en la presente investigación.

10. *Criterios para prácticas evaluativas de calidad* (Grupo GRAPA-RIMA, 2012). Desde su constitución en 2008, el grupo GRAPA-RIMA de la Universidad Politécnica de Catalunya ha realizado varias contribuciones sobre buenas prácticas en materia evaluativa. Fruto de la experiencia acumulada, y con la finalidad de satisfacer las exigencias de transparencia derivadas del EEES, dicho grupo sintetiza ocho criterios para orientar prácticas evaluativas de calidad sobre el aprendizaje universitario. Por cada uno de ellos, incluye una rúbrica con tres niveles de calidad, lo que facilita su aplicación en procesos de planificación o revisión evaluativa. Además de estar ubicada en un contexto próximo, el trabajo del grupo GRAPA-RIMA está fundamentado en antecedentes que también han sido considerados en la presente investigación⁴⁰.

Las conclusiones a las que ha llevado el análisis de cada una de las anteriores referencias, se muestra resumido en la tabla 35. En total, su recorrido ha conducido a la identificación de 19 rasgos o atributos evaluativos, agrupados en siete bloques (A-G). Estos bloques representan las grandes categorías detectadas en el análisis y, por tanto, ofrecen una perspectiva amplia del contenido de los decálogos y manifiestos. Así consta en el resumen de aportaciones declarado en la tabla 34.

³⁹ Se presentan en el bloque siguiente, dentro de los antecedentes que proporcionan propuestas estructuradas.

⁴⁰ Principalmente, las pautas de Nicol (2009) y la referencia proporcionada por el Proyecto REAP (2005).

Tabla 35

Elementos clave para la definición de una teoría de la evaluación: análisis de los antecedentes teóricos

SÍNTESIS: DECÁLOGOS Y MANIFIESTOS	REFERENCIAS									
A. PLANIFICACIÓN ESTRATÉGICA	1	2	3	4	5	6	7	8	9	10
ALINEACIÓN. Coherencia de la evaluación con objetivos de aprendizaje profundos. Formulación de tareas adecuadas, no reproductivas, centradas en los aspectos clave del aprendizaje, que exijan la integración de diversos recursos.		X	X			X	X		X	X
ORIENTACIÓN AL APRENDIZAJE. Diseño de un sistema evaluativo basado en la comprensión profunda del aprendizaje que desea promover y del efecto que la evaluación provoca en el mismo (elección consciente de plazos, procedimientos, etc.).	X	X			X		X		X	
FORMATIVA Y CONTINUA. Integración de la evaluación, que no aparece como una función exclusivamente acreditativa y aislada en la recta final del ciclo formativo. Se planifica desde el comienzo, como parte del proceso de aprendizaje y enseñanza, y se emplea con fines pedagógicos y sumativos. Destaca especialmente el valor pedagógico de la evaluación diagnóstica por ser la menos empleada tradicionalmente.	X						X	X	X	X
CANTIDAD Y EXIGENCIA ADECUADA. Sensibilidad con el proceso de aprendizaje del alumnado. La evaluación no plantea exigencias excesivas que generen indefensión o ansiedad. Favorece la adecuada distribución del trabajo y el esfuerzo que requiere la construcción del aprendizaje deseado.	X	X	X	X		X			X	
ADECUACIÓN Y VARIEDAD DE INSTRUMENTOS Y TAREAS. Empleo de instrumentos apropiados y variados para valorar los aprendizajes pretendidos (características de lo que pretende evaluarse, niveles de complejidad variada, respuesta a la diversidad de estilos y necesidades de aprendizaje, etc.).	X	X						X	X	X
B. PUBLICIDAD Y TRANSPARENCIA	1	2	3	4	5	6	7	8	9	10
INFORMACIÓN CLARA. Transparencia sobre los criterios y condiciones generales que rigen la evaluación del aprendizaje.	X	X	X				X		X	X
EXPECTATIVAS CLARAS. Instrucciones y expectativas claras (tareas, buena ejecución o desempeño, etc.).		X	X	X	X	X	X		X	
INTERIORIZACIÓN. Adopción de medidas para que el alumnado llegue a interiorizar las expectativas y exigencias evaluativas comunicadas.		X			X		X	X		
ALTAS EXPECTATIVAS. Comunicación de altas expectativas de aprendizaje por medio de las exigencias y condiciones evaluativas.			X							
C. COMUNICACIÓN DE RESULTADOS DE APRENDIZAJE (FEEDBACK)	1	2	3	4	5	6	7	8	9	10
FEEDBACK. Devolución de información clara al alumnado sobre el progreso y los resultados de su aprendizaje (comprensible, basada en criterios, unida a objetivos de aprendizaje, puntual, frecuente, seguro, estimulante, etc.).	X	X	X	X	X	X	X	X	X	X
D. ORIENTACIÓN A LA MEJORA	1	2	3	4	5	6	7	8	9	10
REGULACIÓN APRENDIZAJE (PROACCIÓN). Devolución de pautas que orienten la mejora de los resultados de aprendizaje evidenciados y su desarrollo.		X	X	X		X	X	X	X	X

SÍNTESIS: DECÁLOGOS Y MANIFIESTOS	REFERENCIAS									
REGULACIÓN DE LA ENSEÑANZA. Mejora y adecuación de la enseñanza a partir de los resultados puestos de manifiesto en la evaluación.	X	X	X	X			X			
REVISIÓN DE LA EVALUACIÓN. Revisión de los sistemas de evaluación para su mejora continua.	X						X		X	
E. IMPLICACIÓN DEL ALUMNADO	1	2	3	4	5	6	7	8	9	10
PARTICIPACIÓN DISCENTE. Implicación del alumnado en la función evaluativa, de modo que se establezca un proceso de diálogo sobre el aprendizaje a realizar y su evaluación. Esta participación puede tener lugar en el desarrollo del proceso de evaluación (aplicación de formatos participativos) y/o en su propio diseño (negociación de condiciones, plazos, procedimientos, etc.).		X	X	X	X	X	X	X	X	X
F. PROPIEDADES DE LA EVALUACIÓN	1	2	3	4	5	6	7	8	9	10
RIGOR Y FIABILIDAD. Precisión de la evaluación, que discrimina con acierto distintos niveles de logro en el aprendizaje discente. Análisis de su capacidad para devolver una valoración fiable y precisa del aprendizaje y adopción de medidas para contrarrestar los principales riesgos (mecanismos de control, medidas para la reducción del riesgo de plagio, evaluación anónima para reducir sesgo, etc.).	X	X					X	X	X	
VALIDEZ. Valoración real del aprendizaje en toda su globalidad y complejidad. La evaluación se ciñe al aprendizaje que pretendía evaluarse, no aspectos periféricos o relacionados.	X	X			X		X		X	
EQUIDAD. Equidad evaluativa, que no proporciona un trato de favor a determinado sector del alumnado o desfavorecerlo, por la adopción de determinados criterios y condiciones. Tampoco por prejuicios o discriminaciones sujetas a factores de diverso orden.								X	X	
ACORDE A LAS NORMAS Y ESTÁNDARES. Cumplimiento de las normas y estándares evaluativos de referencia en su ámbito de aplicación. Estos pueden ser externos (establecidos a nivel nacional) o internos (normativa y regulaciones propias de la institución).							X		X	
G. COMUNIDADES DE APRENDIZAJE	1	2	3	4	5	6	7	8	9	10
DESARROLLO DE COMUNIDADES DE APRENDIZAJE. Creación de comunidades de aprendizaje basadas en la colaboración y el intercambio por medio de la evaluación, rehusando la competitividad. Pueden involucrar a estudiantes, profesionales (redes de colaboración e intercambio, coordinación, etc.) o ambos colectivos.				X	X				X	

Nota: Las columnas de la derecha recogen los antecedentes consultados en el mismo orden de presentación: (1) Brown et al. (1996); (2) Ramsden (2003); (3) Proyecto REAP (2005); (4) Nicol, 2007, 2009; (5) WMG-ASke, 2007; (6) Gibbs & Simpson, 2009; (7) Boud & asociados, 2010; (8) NUS, 2010; (9) CLE-UB, 2011; (10) Grupo GRAPA-RIMA (2012).

Dada la falta de consenso acerca de los parámetros que deben guiar la evaluación en la universidad, la literatura específica da cuenta de otros precedentes que también han tratado de sintetizar referencias teóricas para su ejercicio. Por su proximidad geográfica y cronológica, resultan de particular interés las conclusiones de cuatro de ellos (ver tabla 36). Las orientaciones contenidas en los mismos confirman, a grandes rasgos, la relevancia de

los elementos clave delimitados en la tabla 35 y reafirman su pertinencia. A su vez, proporcionan una nueva aportación para completar la selección de rasgos evaluativos deseables.

Tabla 36

Rasgos teóricos de la buena evaluación: conclusiones comunicadas en otras propuestas

CRITERIOS Y CONDICIONES DE LA BUENA EVALUACIÓN EN LA UNIVERSIDAD			
Suskie (2006)	Cano et al. (2010)	Buscà, Capllonch et al. (2010)	Cano & Ion (2012)
<i>Rentable, eficiente</i>	Coherente	Principio de adecuación	Potencialidad didáctica
Orientada a objetivos claros	Proporcionada	Principio de relevancia	Valor formativo
Enlazada con decisiones	Formativa	Principio de credibilidad	Coherencia
Valorada, considerada	Continuada	Principio de integración	Variedad de agentes e instrumentos
Fiable, clara, veraz	Variada	Principio de continuidad	Recurso de regulación y autorregulación
	Acreditativa	<i>Principio de viabilidad</i>	
	Diagnóstica	Principio de ética	
	Flexible	Principio de emancipación	
	Contextualizada		
	Individualizada		
	Motivadora		
	Amistosa		
	Rápida		
	Explicitada		
	<i>Viable</i>		
	Precisa		
	Ética		
	Justa		
	Colectiva		
	Meta-evaluativa		

Nota: La fuente de color rojo señala los rasgos que no han sido expresamente señalados en la síntesis de decálogos presentada.

- Suskie (2006), actual directora de la AALHE (Association for the Assessment of Learning in Higher Education), realizó hace algunos años un trabajo de síntesis similar al que está siendo abordado en el presente capítulo. Con el objetivo de identificar los rasgos de la buena evaluación en la educación superior, analizó el contenido de 19 decálogos y manifiestos. Detectó cinco claves o rasgos comunes a todos ellos. Salvada la eficiencia de la evaluación, todos están presentes entre los rasgos evaluativos destacados.
- En el contexto de la Red de Evaluación Formativa y Compartida, Buscà, Capllonch et al. (2010) presentaron ocho principios rectores para guiar el diseño y la puesta en práctica de dispositivos evaluativos. Analizados desde la perspectiva conceptual,

concluimos que la síntesis propuesta ofrece buen reflejo de los mismos. La viabilidad evaluativa representa la única condición que no ha sido expresamente señalada en los decálogos analizados.

- De forma más descriptiva, Cano et al. (2010) elaboraron una síntesis de referencias teóricas en materia evaluativa, llegando a concluir una colección de 20 rasgos característicos de la buena evaluación. En conjunto, devuelven una imagen consistente con la resumida en el presente apartado. Nuevamente la excepción se sitúa en la viabilidad evaluativa.
- La pertinencia de la síntesis efectuada también se respalda en el trabajo de Cano e Ion (2012), quienes establecen cinco condiciones deseables para la práctica evaluativa, afines a las detectadas en la revisión de decálogos presentada.

En este contraste de referencias cabría sumar los 22 consejos que publica el Assessment Resource de la Universidad de Hong Kong (AR@HKU, 2009d) para el diseño de evaluaciones efectivas. Las orientaciones incorporadas en el mismo se sitúan van en línea con todas estas propuestas. Aunque ofrece una visión muy clara e interesante, esta última contribución está formulada desde una perspectiva mucho más práctica y operativa, por eso no ha sido integrada en el banco de decálogos y manifiestos.

1.3.2 Propuestas estructuradas de organismos y centros para la evaluación del aprendizaje universitario

Los trabajos integrados en este segundo bloque también establecen bases para la definición de la evaluación de calidad, pero con un mayor grado de estructuración que los decálogos y manifiestos presentados en el anterior apartado. Las tres nuevas referencias proceden de organismos internacionales y enumeran pautas o directrices para orientar la práctica evaluativa de los centros superiores y universidades situadas dentro de su ámbito de influencia⁴¹. En consecuencia, muestran una perspectiva institucional de la evaluación que excede el ámbito de incumbencia docente y la interacción didáctica al nivel de grupo-aula. A diferencia de las ideas recopiladas en la revisión de decálogos y manifiestos, las aportaciones que se desprenden de estas propuestas se orientan fundamentalmente a las/os responsables institucionales de la evaluación del aprendizaje.

⁴¹ No se incluyen en esta selección las Directrices para la Garantía de Calidad en el Espacio Europeo de Educación Superior (ENQA, 2009) por haber sido integradas en el análisis de los referentes normativos.

1. *Principios para la buena práctica de la evaluación de las/os estudiantes* (Astin et al., 1992- AAHE). Esta primera propuesta se desarrolla a comienzos de la década de los noventa, bajo el auspicio de la Asociación Americana para la Educación Superior. Surge del trabajo conjunto de un grupo de expertas/os de reconocido prestigio internacional, con el propósito de ofrecer un marco de referencia común para la evaluación del aprendizaje en las universidades de su ámbito de influencia. En la primera propuesta se articularon nueve principios fundamentales para orientar su desarrollo; en una revisión posterior, Banta et al. (1996) agregaron un principio más para reflejar la responsabilidad institucional con respecto a la evaluación y la importancia del liderazgo y el apoyo administrativo. El resultado es una colección de principios generales para guiar la evaluación que, dos décadas más tarde y fuera de su contexto originario, sigue teniendo sentido y vigencia.
2. *Indicadores de la evaluación efectiva para las universidades Australianas* (CSHE, 2002). En respuesta a la petición del Comité de enseñanza de universidades australianas, el CSHE presenta en 2002 un proyecto denominado *Evaluación del aprendizaje*. Su objetivo es ofrecer orientación para promover prácticas evaluativas eficaces que refuercen el aprendizaje de las/os estudiantes y se materializa en una página web abierta a todas las universidades del país⁴². Desde ahí, establece tres principios clave a modo de referencia común y compartida y 16 indicadores para favorecer su traslado a la praxis evaluativa. El proyecto incorpora además cinco guías de apoyo para favorecer la consecución de las actuaciones evaluativas previstas en los principios e indicadores detallados.
3. *Código de actuación para el aseguramiento de la calidad académica y los estándares en educación superior* (QAA, 2011). Como se desprende de su título, esta nueva referencia explicita el código de calidad para la educación superior en el Reino Unido. Representa el marco de referencia para todos los agentes implicados en la misma; detalla las expectativas públicas con respecto a la educación superior y las obligaciones a las que deben responder las instituciones formativas. La que se analiza en el presente trabajo representa la tercera versión del código; la primera fue publicada en el año 2000 y se revisó seis años más tarde dando lugar a una segunda edición. En agosto de 2014 acaba de entrar en vigor una cuarta edición del código, sin vocación de ser la final. Este proceso de revisión y adecuación toma respaldo de diversos informes sobre el estado de la educación superior en el Reino

⁴² Su contenido se encuentra disponible en el sitio de Internet del citado comité

<http://www.cshe.unimelb.edu.au/assessinglearning/index.html>

Unido, así como de los *Criterios y Directrices para la Garantía de Calidad en el Espacio Europeo de Educación Superior* (ENQA, 2009)⁴³.

A los efectos de la presente investigación resulta de particular interés el sexto capítulo de la sección B del código (Asegurar y mejorar la calidad académica). Dicho capítulo trata sobre la evaluación de las/os estudiantes y la acreditación de la formación previa. En la edición considerada (QAA, 2011) la propuesta incluye 15 indicadores, organizados en torno a 12 núcleos⁴⁴. Ofrece una referencia para favorecer la buena evaluación en las universidades y encomienda a los propios centros su implementación, control y seguimiento. En consecuencia, se trata de una propuesta con un marcado acento institucional, dirigida a administradoras/es y responsables organizativos, más que al profesorado que va a desarrollar directamente la evaluación.

Se trata de una referencia sólida y ampliamente representada en la literatura, dada su implantación en el contexto universitario del Reino Unido. Además de las reinterpretaciones que se plantean para la aplicación del código en los propios centros formativos⁴⁵, la QAA elabora documentos generales de apoyo para favorecer su desarrollo. Entre ellos, destacamos a modo de ejemplo los documentos para la difusión de buenas prácticas (QAA, 2012a) o los manuales específicos para profesorado novel (QAA, 2012b).

El análisis de estas propuestas aporta a la reflexión una perspectiva institucional que contribuye a dibujar un marco evaluativo más amplio y realista. Sus aportaciones sugieren un mayor grado de divergencia que el detectado en el caso de los decálogos y manifiestos. En suma, las conclusiones extraídas apuntan a nueve categorías centrales en las que se incorpora un elevado número de descriptores o indicadores (ver tabla 37).

⁴³ Esta referencia ha sido considerada al explorar los referentes normativos para la definición del ITC.

⁴⁴ La última edición, publicada pocos meses antes de culminar la presente investigación, reelabora dichos indicadores e incorpora tres más.

⁴⁵ Sirva de ejemplo el realizado por el prestigioso Centre for Excellence in Teaching and Learning in Assessment for Learning de la Northumbria University (<http://www.northumbria.ac.uk/sd/central/ar/teia/enhance/assessfeedback/?view=Standard>) o la elaborada desde la Universidad de York (<http://www.york.ac.uk/media/abouttheuniversity/supportservices/academicregistry/registryservices/guide/Guide%202012-13.pdf>).

Tabla 37

Elementos clave para la definición de una teoría de la evaluación: análisis de las propuestas estructuradas

SÍNTESIS: PROPUESTAS ESTRUCTURADAS	REFERENCIAS		
PLANIFICACIÓN ESTRATÉGICA	1	2	3
ORIENTADA AL APRENDIZAJE. Diseño de sistemas de evaluación ideados conscientemente para motivar el aprendizaje y el desarrollo de los hábitos pretendidos en el alumnado.		X	X
INTEGRADA. Evaluación integrada, desarrollada de forma continua a lo largo del proceso formativo, no final o aislada.	X	X	
FORMATIVA Y SUMATIVA. Combinación de las funciones pedagógicas y sociales de la evaluación.	X	X	X
ALINEACIÓN. Evaluación alineada con los aprendizajes perseguidos. Empleo de sistemas de evaluación centrados en los aprendizajes relevantes y coherentes con el tipo de trabajo que requiere su desarrollo.		X	
EVALUACIÓN COMPLEJA Y CONTEXTUALIZADA. Diseño y desarrollo de tareas diversas que más allá de la mera reproducción, exijan la integración de distintos elementos en situaciones de aplicación y transferencia.	X	X	
CANTIDAD Y EXIGENCIA ADECUADA. La evaluación es sensible con el proceso de aprendizaje del alumnado, sin plantear exigencias excesivas (coordinación de exigencias, planificación de niveles de dificultad crecientes a lo largo de la titulación, etc.).		X	X
PUBLICIDAD Y TRANSPARENCIA	1	2	3
EXPLICITACIÓN. Explicitación de las finalidades de la evaluación, de los aprendizajes que desean promoverse y de las condiciones y criterios aplicables.	X		X
INTERIORIZACIÓN. Adopción de medidas para facilitar la comprensión e interiorización de criterios y condiciones de evaluación por parte del alumnado.			X
COMUNICACIÓN RESULTADOS DE APRENDIZAJE	1	2	3
FEEDBACK. Comisión de retroalimentación o feedback al alumnado sobre sus logros de aprendizaje y su evolución (en diferentes momentos y formatos, información comprensible, frecuente y puntual, etc.).		X	
IMPLICACIÓN DEL ALUMNADO EN LA EVALUACIÓN	1	2	3
EVALUACIÓN PARTICIPATIVA. Participación del alumnado en la evaluación de su aprendizaje (intervención en el diseño, implementación, conclusión o revisión, espacios de optatividad en cuanto a plazos y procedimientos, etc.)	X	X	
PARTICIPACIÓN CONSCIENTE. Participación consciente del alumnado en la evaluación, con conocimiento y consideración de sus derechos y deberes dentro de la misma, y de las normas que la rigen.			X
POLÍTICA INSTITUCIONAL PARA LA EVALUACIÓN DEL APRENDIZAJE	1	2	3
POLÍTICA INSTITUCIONAL. Existencia de una política o normativa (departamental, de facultad o institucional) que guía las prácticas evaluativas individuales y favorece su coordinación. Su delimitación debe tener lugar meditado el efecto de las medidas que plantea en el aprendizaje del alumnado.		X	X
REVISIÓN INSTITUCIONAL DE LA EVALUACIÓN. Establecimiento de cauces y condiciones institucionales propicias para la revisión, la inspección y la mejora de la evaluación del aprendizaje que tiene lugar en su seno.	X		X

EVALUACIÓN PROFESIONAL	1	2	3
COMPETENCIA EVALUATIVA. La institución garantiza que la evaluación se lleve a cabo por docentes con competencia profesional, de modo que queden cubiertas sus funciones ante el alumnado y la sociedad (conocimiento profundo sobre los efectos de la evaluación, normas aplicables, etc.)	X		X
INFORMACIÓN INSTITUCIONAL. Información clara y transparente de la institución al profesorado sobre normas y condiciones que deben satisfacerse en los procesos evaluativos sobre el aprendizaje discente.			X
RENDICIÓN DE CUENTAS INSTITUCIONAL. Rendición de cuentas de la institución sobre los resultados de aprendizaje constatados en los procesos evaluativos internos.			X
PRECISIÓN DE LA EVALUACIÓN	1	2	3
EVALUACIÓN CRITERIAL PRECISA. Adjudicación de calificaciones basada en resultados de aprendizaje claramente delimitados y criterios que discriminan diferentes niveles de logro.		X	X
CONTROL DE AMENAZAS. Adopción de medidas para minimizar el riesgo de plagio.		X	
PUESTA EN PRÁCTICA COHERENTE	1	2	3
CUMPLIMIENTO COHERENTE. Las normas, criterios y compromisos evaluativos declarados en la normativa de referencia encuentran un cumplimiento fiel en la práctica.			X
EVALUACIÓN EQUITATIVA	1	2	3
EVALUACIÓN EQUITATIVA. Delimitación de tareas evaluativas libres de sesgos que pueden situar a determinados grupos de estudiantes en desventaja (lenguas minoritarias, necesidades específicas, etc.).		X	X

Nota: Las columnas de la derecha recogen los antecedentes consultados en el mismo orden de presentación: (1) AAHE (Astin et al., 1992); (2) CSHE; (3) QAA (2011).

El modelo ITC integra en su desarrollo este resumen de rasgos, aunque con un reflejo menos directo por estar dirigidas a las/os responsables institucionales más que al profesorado (tabla 34). Esta perspectiva explica la referencia a aspectos amplios como la política evaluativa general, la evaluación profesional o la puesta en práctica coherente. También el reflejo comedido de aspectos como el feedback o la implicación del alumnado en la evaluación, ampliamente destacados en los decálogos y manifiestos.

1.3.3 Propuestas estandarizadas para la evaluación del aprendizaje

El bloque final de antecedentes incluye dos referencias sustancialmente diferentes de las expuestas hasta el momento: los estándares para la evaluación de los aprendizajes de las/os estudiantes del JCSEE (2007) y los propuestos desde el CRESST (Baker et al., 1996; CRESST, 2012; Land, 1997).

Para responder con mayor precisión a los objetivos de la búsqueda, sólo han sido seleccionadas propuestas de estandarización expresamente referidas a la evaluación del

aprendizaje. Esta decisión excluye del análisis otros intentos de sistematización de la praxis evaluativa efectuados desde una perspectiva más amplia (evaluación institucional, de programas, servicios, etc.). Entre ellos se encuentran los estándares para la evaluación de personal y programas del JCSEE (1998, 1981), las directrices para la evaluación africana dictadas por la African Evaluation Association (AfrEA, 2002) o los Principios Rectores para Evaluadores de la American Evaluation Association (AEA, 2004).

Frente a los decálogos o manifiestos y las propuestas estructuradas, las referencias incluidas en este tercer bloque difieren en el grado de sistematización y en la referencia educativa que adoptan:

- Por una parte, representan propuestas sometidas a un alto grado de rigor y consenso. Establecen estándares para guiar la evaluación, esto es, principios con un amplio respaldo a nivel internacional, a cuyo cumplimiento se asocia una práctica evaluativa profesional de calidad (Baeza et al., 2006; JCSEE, 2007).
- Por otra parte, representan una referencia amplia para la investigación en curso, en la medida en la que no se orientan exclusivamente a la evaluación del aprendizaje en la universidad. En ambos casos, los estándares planteados sirven para guiar la praxis evaluativa en los diversos niveles educativos. No así otras propuestas que, aunque interesantes, han sido descartadas por dedicarse exclusivamente al contexto escolar de la evaluación del aprendizaje discente⁴⁶.

La amplitud de los estándares y su orientación inespecífica a la educación superior, hacen de ellos una influencia indirecta en la delimitación del modelo ITC. Más que aportar contenido directo, los estándares consultados representan el telón de fondo sobre el que se realiza la construcción de la teoría evaluativa de base. Desde esta perspectiva, mientras que el modelo ITC recoge en sí los rasgos propios que debe contemplar la evaluación del aprendizaje universitario, entendemos que los estándares ofrecen una guía general para su implementación, desarrollo o revisión.

Para efectuar la exposición de estos dos últimos referentes se ha sustituido el criterio cronológico por el de amplitud y relevancia:

⁴⁶ Tal es el caso de los *Criterios para la evaluación de los sistemas evaluativos* de Harlen (2007). En su propuesta la autora profundiza en el significado de los criterios determinados para la evaluación en el *Assessment Systems for the Future Project* (ASF), adecuándolos a los diferentes usos que la misma puede adoptar en el trabajo con alumnado menor. Dicho proyecto representa una iniciativa desarrollada entre los años 2003 y 2006 por el Assessment Reform Group (Harlen, 2003).

- *Estándares para la evaluación de los aprendizajes de las/os estudiantes* (The Joint Committee on Standards for Educational Evaluation- JCSEE, 2007). Siguiendo con los esfuerzos del JCSEE por incrementar la calidad en la evaluación de personal (JCSEE, 2001) o de programas (JCSEE, 1998), en 2004⁴⁷ salen a la luz los estándares para la evaluación de los aprendizajes de las/os estudiantes. Bajo la dirección de Gullickson, se propone un sistema de 29 estándares, agrupados en torno a cuatro grandes principios asociados a la calidad evaluativa: adecuación o propiedad, utilidad, viabilidad y precisión (ver tabla 38). Los estándares identifican y promueven la práctica fundamentada de la evaluación de los aprendizajes discentes, favoreciendo el diálogo reflexivo sobre los aspectos clave.

Tabla 38

Propuesta de estándares para la evaluación de los aprendizajes discentes (JCSEE, 2007, pp. 4-5)

ADECUACIÓN. Hace referencia a la legitimidad o propiedad de la evaluación, al desarrollo ético y legal de misma preservando la integridad del sistema y los derechos del alumnado.	
P1	<i>Satisfacción</i>
P2	<i>Políticas y procedimientos apropiados</i>
P3	<i>Acceso a los resultados de la evaluación</i>
P4	<i>Tratamiento del alumnado</i>
P5	<i>Derechos del alumnado</i>
P6	<i>Evaluación equilibrada</i>
P7	<i>Conflictos de interés</i>
UTILIDAD. Hace referencia a la satisfacción efectiva de las necesidades de información de los diferentes sectores involucrados en la evaluación. Se centra en favorecer evaluaciones informativas y oportunas, para que los resultados derivados tengan impacto en la mejora.	
U1	<i>Orientación positiva</i>
U2	<i>Identificación de usuarios y usos</i>
U3	<i>Foco y selección de la información</i>
U4	<i>Credibilidad del evaluador</i>
U5	<i>Identificación de principios</i>
U6	<i>Informes eficaces</i>
U7	<i>Seguimiento</i>

⁴⁷ La edición traducida al español se publica tres años más tarde y es la referida a lo largo de este trabajo.

VIABILIDAD. Hace referencia a la eficiencia del sistema evaluativo, a los recursos requeridos para su eficaz desarrollo y al compromiso y participación de los diferentes estamentos implicados.

V1 *Procedimientos prácticos*

V2 *Viabilidad política*

V3 *Apoyo a la evaluación*

PRECISIÓN. Hace referencia a la adecuación técnica de la información revelada y transmitida a partir de la evaluación.

PR1 *Procedimientos de evaluación claros*

PR2 *Análisis del contexto*

PR3 *Procedimientos descritos*

PR4 *Fuentes de información defendibles*

PR5 *Uso válido de la información*

PR6 *Información fiable*

PR7 *Manejo de la información y control de calidad*

PR8 *Análisis de información cuantitativa*

PR9 *Análisis de información cualitativa*

PR10 *Conclusiones justificadas*

PR11 *Control del sesgo*

PR12 *Metaevaluación*

Estos estándares son de interés para cualquiera de las audiencias implicadas en el proceso evaluativo, aunque se orientan fundamentalmente al colectivo docente por su especial implicación. A pesar de la sistematización que acompaña a su definición, los estándares no son prescripciones, ni tampoco prevén un determinado tipo de respuesta para cada tipo de demanda o situación. En suma, representan un recurso para diseñar y valorar la evaluación de los aprendizajes del alumnado desde la reflexión y la elección consciente. Así lo han constatado Cabra (2008) y C. L. Rodríguez (2008) al investigar sobre la aplicación de estos estándares en el contexto específico de la educación secundaria.

- *Requisitos para la evaluación del aprendizaje del alumnado (CRESST).* El Centro Nacional de Investigación sobre Evaluación, Estándares y Medición del Estudiante es un centro dedicado a la promoción de la evaluación de calidad, frente a los persistentes problemas constatados en su diseño y desarrollo. Adscrito a la Universidad de UCLA, comienza su andadura en 1966 como Centro para el Estudio de la Evaluación (CSE- *Center for the Study of Evaluation*). Desde entonces, desarrolla una extensa tarea de investigación para contribuir a la calidad de la praxis evaluativa en distintos contextos formativos, incluido el universitario.

En respuesta a sus propósitos, desde el centro se publican diversas contribuciones, entre las que destacamos tres por su particular interés a los fines de la presente investigación. La primera es la firmada por Baker et al. (1996), que presenta los cuatro requisitos a los que debe ajustarse la evaluación para superar las clásicas deficiencias que le acompañan en los distintos niveles educativos: validez, credibilidad, justicia y utilidad. Para procurar su consecución, Land (1997) profundiza y desglosa dichos requisitos en una segunda aportación relevante. La tercera contribución se aproxima más a la perspectiva de los decálogos que a la de los estándares por su amplitud, aunque ha sido mantenida en este bloque por su unión con las dos anteriores; ofrece la enumeración de una serie de retos o líneas de avance para guiar la investigación evaluativa (CRESST, 2012).

El resto de informes de investigación y publicaciones del CRESST consultadas muestran el desarrollo y adecuación de estos requisitos y prioridades en el contexto escolar⁴⁸.

2. Conclusiones: un modelo de evaluación del aprendizaje hacia el que orientar la propuesta de competencias evaluativas específicas

Adoptando la definición que proponen B. García et al. (2008), entendemos por modelo la visión sintética de enfoques o teorías pedagógicas que orienten a especialistas y profesorado en la elaboración y análisis de un objeto en particular, en nuestro caso, la evaluación del aprendizaje universitario. Los modelos representan...

patrones conceptuales que permiten esquematizar de forma clara y sintética las partes y los elementos que se desean observar, así como los componentes de una de sus partes; también se considerará como representación arquetípica o ejemplares de procesos, en los que se exhiben las distribuciones de funciones y de secuencias de forma ideal. (p. 103).

⁴⁸ Sirva de ejemplo la reciente publicación de los *Criterios para una Evaluación de Alta Calidad* (Darling-Hammond et al., 2013), que establece condiciones a las que debe tender la praxis evaluativa en el trabajo con alumnado de Educación Primaria y Secundaria.

Desde esta perspectiva, el modelo ITC representa la teoría de la evaluación a la que se orienta la investigación en curso. Detalla los rasgos deseables que debe presentar la evaluación del aprendizaje en la formación básica de maestras/os, atendiendo a los referentes seleccionados para su configuración (ver figura 18). No existe una única visión sobre el sentido y alcance de la evaluación (Buendía & López, 1999; Cabra, 2008; Porto, 2010) y la calidad es una atribución, no un rasgo incorporado a prácticas y objetos (Tejedor, 2003). En consecuencia, este modelo expone una manera concreta de entender la praxis evaluativa que no es universal ni indiscutible, sin embargo, la propuesta de sistematización de la competencia docente presentada en el próximo capítulo se orienta a ella.

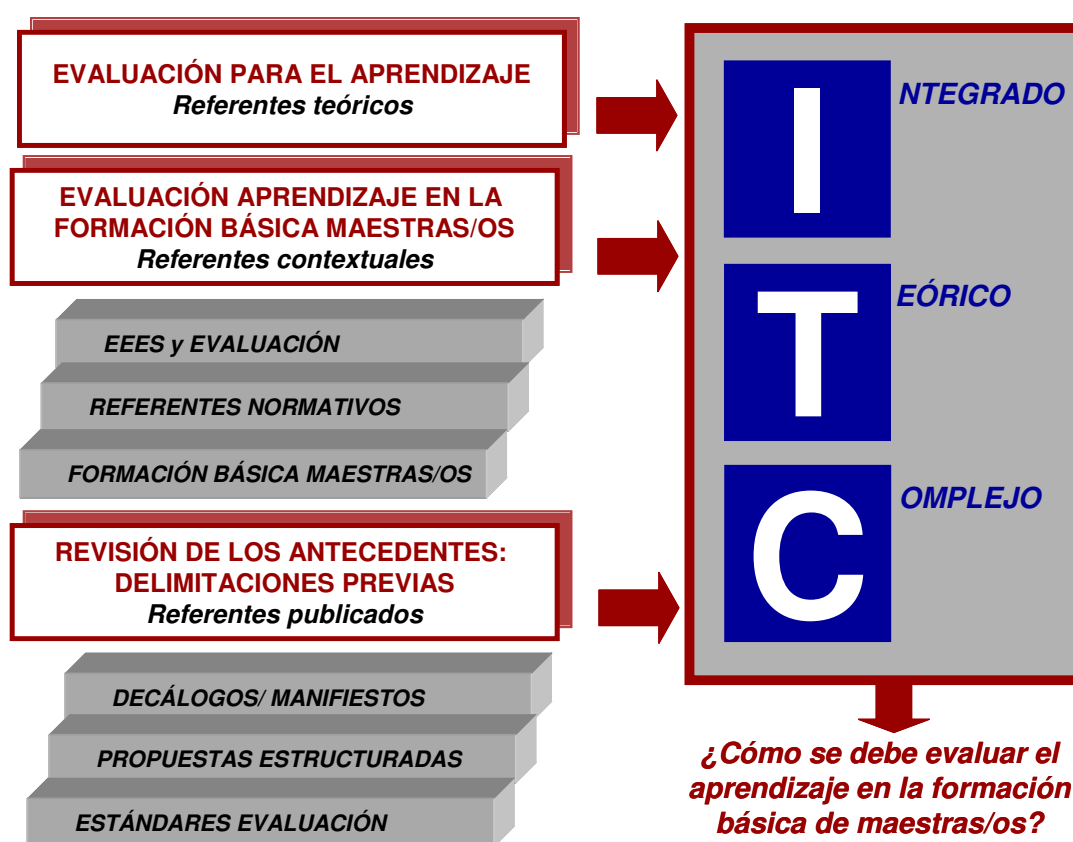


Figura 18. Modelo ITC (Integrado Teórico Complejo) de evaluación.

La propuesta recibe la denominación de **modelo ITC de evaluación del aprendizaje** en referencia a sus principales rasgos distintivos:

- En primer lugar, el modelo presentado es **integrado**, en la medida en que recopila, aúna y fusiona en sí la consideración de todo el repertorio de referentes recorrido en la primera parte del capítulo (orientaciones teóricas, condiciones contextuales y antecedentes).

- En segundo lugar, el modelo ideado plantea una reflexión sobre lo ideal, no sobre requisitos mínimos. Desde esta perspectiva se puede caracterizar como una propuesta **teórica**, que establece los rasgos evaluativos deseables y aporta una dirección clara hacia la que aproximar las prácticas articuladas en este terreno. Representa un marco de referencia para desarrollar el diseño o la revisión de la evaluación del aprendizaje. Su uso puede realizarse de forma global o atendiendo a secciones específicas según necesidades y prioridades particulares.
- Por último, el modelo presentado se caracteriza por su **complejidad**, rasgo connatural al fenómeno que delimita. En consecuencia, proporciona un marco de referencia amplio y extenso que delimita subprocesos, criterios e indicadores de forma ordenada. Recurrir a propuestas más simples y reducidas ofrecería ventajas prácticas evidentes, aunque también el riesgo de incurrir en una artificial simplificación. La mejora de la práctica evaluativa exige el reconocimiento de su complejidad (Santos, 2010a; Zeichner, 2010). Por lo tanto, apostamos por la definición de una propuesta compleja, en cuya aplicación caben aproximaciones parciales, según las necesidades y prioridades particulares.

Partiendo de estas premisas, el modelo ITC detalla en tres niveles de desarrollo progresivo la teoría evaluativa que representa:

- En primer lugar, el modelo identifica siete **subprocesos** implicados en el complejo proceso de la evaluación del aprendizaje discente (ver tabla 39). Estrechamente interrelacionados, cada uno de ellos hace referencia a un espacio de la evaluación e implica la adopción de toda una serie de decisiones y actuaciones con impacto en el conjunto.
- En segundo lugar, en cada uno de los subprocesos o espacios de decisión y actuación, se formulan **criterios**. Estos criterios señalan los requisitos deseables que debe reunir la evaluación del aprendizaje en cada uno de los espacios delimitados. Expresan las características que se esperan de un proceso, producto o servicio (Tejedor, 2003), en este caso, la evaluación del aprendizaje en la formación básica de maestras/os. Reciben la denominación de criterio, dado que la provisionalidad tentativa de la propuesta planteada no se corresponde con el consenso y la fundamentación propia de los estándares. Aun así, pretenden constituir una guía para promover la práctica sensata y razonable de la evaluación de los aprendizajes del alumnado, tal y como hacen los estándares.

- Por último, por cada uno de los criterios definidos, el modelo se completa con la formulación de **indicadores**. Dichos indicadores señalan rasgos que incorporan las evaluaciones cuando satisfacen el criterio al que hacen referencia. Son las medidas necesarias para verificar la presencia o cumplimiento de los criterios predefinidos (Tejedor, 2003). Representan el más flexible de los tres niveles de desarrollo en dos sentidos. Por una parte, constituyen una colección abierta de posibles manifestaciones; incorpora los indicadores más representativos para cada criterio, admitido que pueden existir otros posibles. Por la otra, determinados indicadores se caracterizan como rasgo de criterios diferentes, dada su interrelación teórica y práctica.

La revisión de los referentes detallados en la primera parte del capítulo ha llevado a identificar siete subprocesos desglosados en un total de 19 criterios y 68⁴⁹ indicadores estrechamente relacionados entre sí. La tabla 39 resume los dos primeros niveles de definición y el **ANEXO 04** ofrece una visión completa del modelo, en la que también se incorpora el tercero.

Tabla 39
Modelo ITC: subprocesos y criterios

SUBPROCESO		CRITERIOS	
1	Planificación estratégica del proceso de evaluación	1.1	Potencialidad estratégica
		1.2	Alineación y coherencia
		1.3	Selección metodológica justificada
		1.4	Respuesta a la diversidad
		1.5	Aprendizaje de competencias
		1.6	Evaluación compartida o participativa
2	Comunicación y difusión de criterios y condiciones	2.1	Explicitación y difusión de criterios y condiciones
		2.2	Comprensión e interiorización condiciones básicas
3	Recogidas y registro de evidencias evaluativas	3.1	Implementación ajustada
		3.2	Recopilación y registro personalizado de evidencias de aprendizaje
4	Análisis de resultados	4.1	Control de sesgos
		4.2	Proceso de calificación

⁴⁹ Dada su interrelación, 11 de ellos se repiten en criterios diferentes.

SUBPROCESO		CRITERIOS	
5	Feedback y proacción	5.1	Comunicación de resultados
		5.2	Retroalimentación o feedback
		5.3	Proacción para la mejora del aprendizaje
		5.4	Proacción para la mejora de la enseñanza
6	Revisión del proceso de evaluación (metaevaluación)	6.1	Revisión específica
		6.2	Proacción para la mejora de la evaluación
7	Modelado en la formación básica de maestros	7.1	Modelado deliberado de experiencias evaluativas significativas

Nota: El último subproceso aparece sombreado para marcar su vinculación específica al ámbito de la formación básica de maestras/os.

El tercer apartado del capítulo expone sintéticamente el desarrollo completo del modelo ITC. Antes de ellos, los próximos apartados recorren cada uno de los subprocesos que lo configuran para fundamentar su pertinencia y relevancia. Ésta viene determinada por el respaldo que ofrecen los distintos referentes consultados en su delimitación, así como por las contribuciones derivadas de la literatura e investigación específicas.

2.1 Planificación estratégica del proceso de evaluación.

La evaluación del aprendizaje discente es un proceso sistemático que no debe ser conducido por impulso de la improvisación (Stufflebeam, 1974/2011, 2001b). La planificación representa la primera fase del proceso evaluativo y conlleva la adopción de toda una serie de decisiones que condicionan su efectividad en los usos sociales y/o pedagógicos para los que haya sido concebido.

El modelo ITC establece que el diseño de los dispositivos evaluativos de calidad implica la concurrencia de seis criterios clave.

2.1.1 Potencialidad estratégica de la evaluación.

Adscrito al enfoque teórico de la evaluación para el aprendizaje, el modelo ITC propone la configuración de dispositivos evaluativos centrados en el aprendizaje e ideados para su estimulación consciente. Su potencialidad didáctica reside en el efecto regulador

que naturalmente provoca la evaluación sobre el aprendizaje (Barberá, 1999; Bilbao, 2012; Brown, 2003b; J. Chappuis et al., 2012; 2002; Gibbs & Dunbar-Goddet, 2007; Jessop et al., 2014; Shepard, 2000; Gibbs & Simpson, 2009). De acuerdo con Biggs (2005) dicho efecto opera siempre, exista o no premeditación por parte del profesorado. En consecuencia, resulta indispensable planificar estratégicamente la evaluación para que se ponga al servicio de los aprendizajes previstos en los planes de estudio.

El diseño estratégico de la evaluación exige que el profesorado disponga de una comprensión razonable del proceso de construcción que subyace al aprendizaje (Bain, 2007; Brown et al., 1996). También una cierta cualificación didáctico-pedagógica que le capacite para emplear deliberadamente la evaluación, adoptando decisiones conscientes sobre los efectos de la misma en el aprendizaje discente (J. Chappuis, 2014). Estas decisiones deberían partir de una serie de consideraciones básicas:

- Para que pueda ser estratégicamente empleada, la evaluación debe estar *plenamente integrada* en el proceso de enseñanza-aprendizaje. Fijados con claridad los propósitos formativos, ésta pasa a concebirse como recurso para favorecer su consecución de principio a fin del proceso. Suscrita de forma generalizada, las declaraciones en favor de esta premisa se suceden en publicaciones nacionales e internacionales de las tres últimas décadas (Astin et al., 1992; Black & Wiliam, 1998a; Birenbaum, 1996; Boud & asociados, 2010; S. Chappuis & Stiggins, 2002; Dixon & Rawling, 1987; Gibbs, 2006; Sanmartí & Jorba, 1995; NUS, 2010; Stobart, 2010). Su materialización tiene lugar por medio de la *evaluación continua*, rasgo evaluativo impulsado desde los referentes normativos (Real Decreto 1791/2010) y reforzado con la consolidación progresiva del EEES (De Miguel, 2010; Sabaté et al., 2008). El seguimiento que requiere el desarrollo de las competencias (Cano, 2007, 2011; Poblete, 2007) de un lado, y la adopción del trabajo del alumnado como unidad de referencia en el sistema ECTS (Coronel, 2007) del otro, así lo propician. También los resultados reportados en estudios empíricos desarrollados sobre el tema, que reportan una mejora del rendimiento discente cuando se emplean dispositivos de evaluación continua (Arribas, 2012; Castejón et al., 2012; Fraile et al., 2013; Holmes, 2015; R. Romero et al., 2014).
- Sin embargo, la integración de la evaluación en el proceso formativo debe diseñarse *de forma sensible*, calculando el impacto de las demandas planteadas en el proceso de construcción del aprendizaje. Las/os estudiantes modulan y gradúan su esfuerzo y su trabajo de acuerdo a las exigencias evaluativas percibidas (Gibbs & Simpson, 2009). En consecuencia, el profesorado debe ser capaz de *procurar el esfuerzo y el trabajo requerido para la consecución de los objetivos por medio de la evaluación*

(Chickering & Gamson, 1987, 1999; Haigh, 2007; Torre, 2008). La calidad de la evaluación no es cuestión de cantidad (Cano, 2014; S. Chappuis et al. 2009; Hattie & Timperley, 2007). Exigencias excesivas, descoordinadas o poco ajustadas pueden incidir negativamente en el aprendizaje (Brown et al., 1996; Monereo, 2003; Nicol, 2007, 2009) y en la motivación (ARG, 2002; Harlen & Deakin, 2003; J. Chappuis et al., 2012). También las prácticas evaluativas que hacen de la continuidad una secuencia de demandas acreditativas (Monereo & Castelló, 2009b; Murphy, 2006; Rizo, 2004) y que privan al alumnado del valor formativo del error (J. Chappuis, 2014; Sanmartí, 2007; Torre, 2008; Villa, 2004):

Debería preocuparnos la aparición de falsas profecías que, bajo el nombre de evaluación continua y formativa, constituyen parte de nuevas propuestas curriculares, pero que en el fondo consisten únicamente en calificaciones parciales, fragmentadas en el tiempo; en una distribución de puntos que el profesor suma como si fuera una caja registradora que en el momento del acta emite la cuenta final. (Margalef, 2009, p. 91).

- Para evitarlo, el diseño evaluativo debe prever el empleo estratégico de las funciones pedagógicas, incorporando *demandas formadoras y/o formativas junto a las clásicas pruebas sumativas*. Aunque esta necesidad representa un acuerdo unánime entre los distintos tipos de referente considerados, sigue constituyendo un reto de actualidad según manifiestan las revisiones de I. Álvarez (2008b) o Black y Wiliam (2009) sobre el tema. Así lo evidencia también el planteamiento pedagógico del EEES, que propone para la evaluación un cambio de lógica en este mismo sentido (A. Calatayud, 2008; De Miguel, 2010; M. J. García, 2010; Monescillo & Toscano, 2007).
- El empleo estratégico de la evaluación exige, a su vez, la comunicación clara y explícita de los fines que persigue y los criterios sobre los que reposa. Esta condición da lugar al segundo subproceso propuesto para la evaluación y se aborda en un próximo apartado (*subproceso de la comunicación y difusión de criterios y condiciones*).
- Además de explícitos, los criterios y condiciones definidos para la evaluación deben configurarse en coherencia absoluta a los objetivos de aprendizaje a los que sirven y las estrategias metodológicas previstas para su consecución. Esta premisa remite a la alineación, principio evaluativo referido en el siguiente criterio del subproceso.

2.1.2 Evaluación alineada.

La alineación o coherencia con los objetivos formativos y la metodología dispuesta para su logro es una consecuencia de centrar la evaluación en el aprendizaje discente. Este segundo requisito impone que la evaluación se diseñe en vinculación directa con dos componentes curriculares: las competencias u objetivos formativos previstos⁵⁰ y las actividades dispuestas para alcanzarlos (Biggs, 2005; Biggs & Tang, 2011). Se trata de una condición ampliamente suscrita en la literatura específica (AR@HKU, 2009c; Boud & asociados, 2010; Combs et al., 2008; Grupo GRAMA-RIMA, 2012; Leal, 2004; L. Prieto, 2004).

De acuerdo con J. Chappuis et al. (2012), asegurar este nivel de coherencia exige comprender los resultados de aprendizaje establecidos y el efecto potencial de los métodos de evaluación aplicables en los mismos.

- Para el alumnado las exigencias evaluativas representan, mejor que los objetivos, los retos a superar durante el periodo formativo (Barberá, 1999; Jorba & Sanmartí, 2000; Ramsden, 1991; Sanmartí, 2007). Dependiendo de los requisitos y condiciones que establezca la evaluación, se potenciará en el alumnado un determinado tipo de operación mental y enfoque de trabajo (Joughin, 2003; Santos, 2003) y, con ello, una mayor o menor profundidad y calidad en el aprendizaje resultante (Biggs, 2005; Sans, 2008). Por ese motivo, la evaluación no debe desviarse de los objetivos previstos en los programas y planes de estudio; su diseño debe partir de un análisis pormenorizado de los propósitos formativos, que posibilite decisiones evaluativas coherentes. Este requisito resulta particularmente relevante en las profesiones reguladas como la de maestra/o, en las que median implicaciones normativas.
- La alineación de la evaluación con la metodología lleva a considerar las tareas evaluativas tareas de aprendizaje (Carless, 2007, 2009; Carless, Joughin, Liu, & asociados, 2006). La integración de la evaluación en la planificación del proceso de enseñanza- aprendizaje difumina los límites entre las actividades de aprendizaje y evaluación (Coll et al., 2007; Yániz & Villardón, 2006) y entre enseñanza, aprendizaje y evaluación (Bernabé, 2008; Bezanilla et al., 2014). “Se puede afirmar que las actividades evaluativas no son sino actividades educativas y la distinción entre unas y otras es puramente metodológica o académica [...]” (Mateo, 2006a, p.

⁵⁰ Entendiendo que los planes de estudio de las titulaciones incorporan competencias genéricas y específicas, para cuyo desarrollo pueden formularse objetivos formativos intermedios.

196). En consecuencia, el diseño de tareas de evaluación o aprendizaje pasa a constituir un elemento clave del diseño (Knight, 2005). Así lo contempla el próximo criterio, que profundiza sobre la elección consciente y justificada que subyace a dicha decisión.

Junto a este nivel de alineación interna dentro de un módulo o asignatura determinada, la evaluación también debería atender a un nivel más amplio de coherencia con respecto al planteamiento general de la titulación o el plan de estudios.

En el marco universitario, la práctica de evaluación no puede continuar teniendo como referente la asignatura y el profesorado (considerado individualmente), sino que se tiene que considerar el conjunto de asignaturas y, por lo tanto, el equipo docente tanto desde una perspectiva transversal (qué competencias se trabajan y se evalúan el primer trimestre, por ejemplo) como longitudinal (de qué manera las diferentes asignaturas contribuyen en diferentes niveles al desarrollo de una competencia). (Gairín, 2009, p. 28).

Este segundo nivel de alineación exige que la evaluación se diseñe y desarrolle de manera coordinada en el equipo docente (Briggs et al., 2003; Cano et al., 2010; Sayós, 2010). Además de evitar prácticas desalineadas con los logros generales asociados a la titulación, esta coordinación puede contribuir a la disminución de la sobrecarga de tareas evaluativas para estudiantes y profesorado.

2.1.3 Selección metodológica justificada.

El poder de influencia reconocido para la evaluación hace que los instrumentos y tareas que se empleen predispongan la atención del alumnado en una u otra dirección (Castejón et al., 2009; Monroy & Hernández, 2014; Navaridas, 2002; Scouller, 1996, 1998; Struyven et al., 2005). La metodología de evaluación no es neutra y, en consecuencia, su elección debe realizarse de forma consciente y justificada (Arribas, 2012).

Frente al protagonismo que tradicionalmente se ha otorgado a los instrumentos evaluativos (Verdejo, 2008), en el modelo ITC su elección no representa sino una decisión más de todas las que articulan el diseño de un sistema evaluativo centrado en el

aprendizaje. Para favorecer que así sea, la planificación de las herramientas y tareas debe realizarse procurando niveles complementarios de adecuación:

- *Adecuación pedagógica.* En directa relación con el criterio de la alineación evaluativa, las tareas y herramientas dispuestas para la recogida de evidencias deben responder a los aprendizajes previstos. Como indica Tejada (2005) “la selección y uso de instrumentos de evaluación está relacionada con qué y cuánta evidencia es suficiente para evaluar la competencia” (pp. 25-26). Esa elección debe procurar que los aspectos centrales del aprendizaje se prioricen y refuercen frente a los elementos secundarios (Trillo, 2005). También debe permitir captar la amplia variedad de manifestaciones a las que éste dé lugar, por lo que se recomienda la combinación de procedimientos diferentes (Brown, 2003a; Brown et al., NUS, 2010).
- *Adecuación práctica.* Las tareas y situaciones evaluativas deben ser adecuadas a las condiciones prácticas previstas para su aplicación (AR@HKU, 2009d; Cano et al., 2010; CRESST, 2012; JCSEE, 2007).
- *Adecuación al perfil discente.* Las herramientas y situaciones dispuestas al servicio de la evaluación deben ser adecuadas para las/os estudiantes. Deben plantear demandas claras y comprensibles de acuerdo a su nivel de desarrollo y estar libres de ambigüedad desde el punto de vista comunicativo (Buscà, Capllonch et al., 2010; Coronel, 2007; Knight, 2005). Su efectividad será mayor en la medida en la que resulten estimulantes y desafiantes para el alumnado (Biggs, 2005; Shepard, 2000; Trillo, 2005).
- *Adecuación metodológica.* Sin perjuicio de lo anterior, los instrumentos y herramientas dispuestas al servicio de la evaluación deben responder a los principios básicos de validez y fiabilidad (ver *subproceso cuarto de análisis de resultados*). En consecuencia, debe procurarse la selección de instrumentos y situaciones que realmente respondan al aprendizaje a valorar y con capacidad de determinar con precisión niveles de logro diferentes respecto al mismo (Boud & asociados, 2010; CLE-UB, 2011; Suskie, 2006).

2.1.4 Respuesta a la diversidad.

La planificación de la evaluación debe llevarse a cabo procurando el respeto a la diversidad natural de talentos y estilos de aprendizaje presentes en el grupo de estudiantes (Chickering & Gamson, 1987, 1999). Frente a las prácticas evaluativas homogéneas y

estereotipadas, los referentes consultados abogan por dispositivos evaluativos capaces de responder a la variedad de necesidades y particularidades entre el alumnado (Brown et al., 1996; Cano et al., 2010; Grupo GRAMA-RIMA, 2012; NUS, 2010), especialmente evidente en la formación universitaria. Esta demanda clásica vuelve a cobrar relevancia con el planteamiento formativo europeísta, donde la diversificación del perfil discente y de las trayectorias formativas acarrea una mayor exigencia de flexibilidad en la evaluación (Rué, 2007). J. Sánchez (2011) afirma que, incidir en el desarrollo de competencias lleva a “contemplar y conjugar distintos tipos de evaluación atendiendo a momentos igualmente diferentes, así como a un alumnado diverso” (p. 41), mientras que Tejedor (2009) señala la necesidad de diversificar las pruebas.

La atención a la diversidad del alumnado desde los mecanismos llamados a valorar el aprendizaje representa una de las cuestiones que mejor explica la complejidad de la evaluación según Coll et al. (2000). Su abordaje puede conjugar intervenciones complementarias:

- Los espacios deliberadamente dispuestos para la flexibilidad y la optatividad representa una alternativa para favorecer la adecuación a las necesidades individuales (Bevitt, 2015; CSHE, 2002; V. M. López, 2008; Navarro, 2010).
- De modo similar, la variedad metodológica representa una medida que, indirectamente, contribuye a responder a estilos de aprendizaje y necesidades diversas entre el alumnado.
- La diversificación también puede ser el resultado de procesos de negociación y consenso con el alumnado (Batalloso, 2004; Capllonch & Buscà, 2012; Monereo & Lemus, 2010). El modelo ITC se orienta a la formación básica de maestras/os de Educación Infantil y Primaria, en el marco de las profesiones reguladas. En consecuencia, abogamos por que esta negociación afecte únicamente a las condiciones de la evaluación, en contra de los testimonios que apoyan la conveniencia de que se extienda a la delimitación de los propios criterios evaluativos (Biggs, 2005; Falchikov & Goldfinch, 2000).

Ante la probada incidencia de los valores e ideologías socioculturales en la evaluación, este requisito también exige adoptar un enfoque inclusivo y comprensivo con respecto a la realidad multicultural (NUS, 2010; QAA, 2011, 2014). En definitiva, “la evaluación ha de romper los moldes de la homogeneidad y uniformidad, propias del modelo productivo, para ofrecer modelos adaptativos y polivalentes, más coherentes con una visión diferenciada e inclusiva” (Bordas & Cabrera, 2001, p. 32).

2.1.5 Evaluación de competencias.

En el contexto específico del EEES, diseñar dispositivos evaluativos centrados en el aprendizaje implica dirigir la evaluación a captar y promover el desarrollo de competencias. Los tradicionales modelos de evaluación están concebidos para evaluar conocimientos teóricos principalmente, por lo que se precisan nuevos planteamientos capaces de responder al nuevo objeto (Alsina, 2011; Arnau, 2009; Bezanilla et al., 2014; Guerrero et al., 2013; Rué, 2007; Yániz, 2008).

“Aunque la necesidad de repensar la evaluación en dirección a la evaluación de competencias está clara, no está tan claro los rasgos en los que debería concentrarse para lograrlo” (Baartman et al., 2007, p. 117). Identificarlos, exige delimitar los atributos de las competencias como objeto de evaluación específico y a ello responde la tabla 40, que expone las características particulares que les brindan identidad propia frente a los contenidos, referencia evaluativa tradicional.

Tabla 40

Características básicas de las competencias como objeto evaluativo específico

RASGO CARACTERÍSTICO	DESCRIPCIÓN
Composición compleja e integrada	Las competencias son el resultado de la movilización e integración armónica de recursos de diversa naturaleza: conceptuales, procedimentales, y actitudinales. Su desarrollo no puede plantearse desde una estrategia aditiva que vaya incorporando elementos aislados, ni debe orientarse a la producción de respuestas mecánicas.
Multidimensionalidad	Producto de una combinación dinámica y compleja, las competencias pueden dar lugar a múltiples manifestaciones. Su desarrollo no debe plantearse como una lista cerrada de componentes a dominar, ni se pueden anticipar todas sus aplicaciones.
Versatilidad	Las competencias representan un recurso versátil para quien las posee; son aplicables en situaciones y contextos diferentes a aquel en el que han sido desarrolladas y con distintos propósitos.
Carácter práctico	Las competencias se desarrollan y manifiestan en la acción o puesta en práctica. Su desarrollo puede entrañar el trabajo sobre saberes o componentes actitudinales, pero el fin último del mismo es la articulación conjunta de los diversos elementos en la ejecución práctica.
Carácter situacional e interactivo	Las competencias se activan y manifiestan en respuesta a situaciones complejas (demandas, resolución de problemas, proyectos o tareas). Tales circunstancias se producen en contextos diferentes y con la implicación de agentes diversos, lo que condiciona posibilidades diversas de realización.
Consciencia y voluntad	Las competencias no representan talentos o habilidades innatas o inconscientes. Se trabajan y se desarrollan, aunque puedan existir predisposiciones personales que faciliten u obstaculicen dicho proceso.

RASGO CARACTERÍSTICO	DESCRIPCIÓN
Carácter reflexivo y reconstructivo	Las competencias no constituyen un logro plenamente abarcable. Se desarrollan y se recrean dependiendo de las experiencias y aprendizajes vivenciados en los espacios educativos formales, informales o no formales.
Carácter cualificado	Las competencias no representan logros mínimos. El calificativo “competente” acompaña a actuaciones cualificadas que responden con solvencia a las circunstancias que las motivan.

Nota: Reelaboración propia a partir de varias contribuciones (Bonsón, 2009; Bonsón & Benito, 2005; Cano, 2007, 2011; Cano & Ion, 2012; Cano et al., 2010; J. M. García, 2010; Le Boterf, 2001; Navío, 2005; Pavié, 2011; Tobón, 2010; Unigarro & Rondón, 2005; Villa & Poblete, 2007, 2011; Viso, 2010).

Atendiendo a las particularidades del marco universitario y a la especificidad de las competencias como objeto de estudio diferenciado de los contenidos, el modelo ITC incorpora nueve criterios para orientar su evaluación. Su aplicación daría lugar a la evolución que ilustra la tabla 41 con respecto a la evaluación de contenidos:

Tabla 41

Implicaciones de la ampliación del objeto de evaluación en el EEES: resumen comparativo (Bilbao, 2011b, pp. 115-116)

	ENFOQUE EVALUATIVO BASADO EN...	
	CONTENIDOS	COMPETENCIAS
OBJETO	Contenidos, básicamente conceptuales.	Movilización de recursos de distinto orden (conocimientos, procedimientos, actitudes) en situaciones concretas.
PERSPECTIVA	Disciplinar, parcializada.	Interdisciplinar, globalizada.
CONTEXTO EVALUATIVO	Teórico. Evaluación descontextualizada.	Real o verosímil. Evaluación contextualizada.
TAREAS EVALUATIVAS	Básicamente teóricas y reproductivas.	Diversas, orientadas a captar la aplicación y transferencia.
INSTRUMENTOS	Por lo general, evaluación monometódica por medio de pruebas a libro cerrado.	Diversidad de instrumentos, en condiciones variables.
FLEXIBILIDAD	Respuestas uniformes. Evaluación convergente.	Respuestas diversas. Evaluación divergente.
COMPLEJIDAD	Baja, basada en conductas y pruebas de baja inferencia.	Alta, basada en respuestas complejas de alta inferencia.
REGISTRO	Simple. Informe final en formato estándar (acta de evaluación).	Complejo. Registro individual de evidencias de logro y avance a lo largo de todo el proceso.
INTERVENCIÓN	Evaluación autocrática en manos del profesorado.	Evaluación participativa con fórmulas variadas de intervención (alumnado, profesorado, otras audiencias).
ÉNFASIS	Enseñanza del profesorado.	Aprendizaje del estudiantado.

	ENFOQUE EVALUATIVO BASADO EN...	
	CONTENIDOS	COMPETENCIAS
RESULTADOS ESPERADOS	Reproducción del contenido de la enseñanza.	Aplicación y transferencia de las competencias incluidas en el perfil profesional delimitado.

Los seis subapartados presentados a continuación resumen los principales rasgos evaluativos que se desprenden de esta evolución.

2.1.5.1 *Evaluación globalizada.*

Las competencias son el resultado de experiencias integradoras de aprendizaje (Voorhees, 2001) y deben ser captadas mediante planteamientos globalizados y holísticos. Para ello, es preciso superar dos limitaciones clásicas:

- Las competencias recogen en sí mismas el saber, el hacer y el ser (J. Contreras, 2010), de modo que la evaluación en la universidad no puede limitarse a uno sólo de los componentes como tradicionalmente ha hecho (Mateo, 2006b; Villardón, 2006).
- El enfoque curricular basado en competencias amplía el área de significado de lo que se entiende por aprendizaje (Hernández et al., 2009; M. Martínez & Viader, 2008; B. Salinas & Cotillas, 2007). En consecuencia la evaluación debe tender a formatos holísticos e interdisciplinarios que superen las clásicas fronteras entre módulos, materias o asignaturas (Bolarín et al., 2013; Cano et al., 2010; Coronel, 2007; Sayós, 2010; Tobón, 2010). Más que la mera concurrencia, la evaluación debe captar la movilización estratégica e integrada de los componentes enumerados ante situaciones y circunstancias diversas, no sólo en el contexto académico, sino también fuera de él (R. McDonald et al., 2000; Rué, 2008).

2.1.5.2 *Evaluación continua y necesidad de seguimiento.*

Las competencias se desarrollan progresivamente y ese rasgo debe reflejarse en su valoración (Gairín, 2009). Se crean y recrean en un proceso reconstructivo (Cano, 2007; 2011; Pavié, 2011), y ello exige realizar una evaluación continua sobre las mismas (Poblete, 2007; Sabaté et al., 2008). Tardif (2006; cit. en A. Fernández, 2010) caracteriza este cambio como el paso del planteamiento fotográfico de la evaluación, centrado en el juicio final, al planteamiento videográfico, en el que se procura la progresión del desarrollo de

competencias. Como será detallado al presentar el tercer subproceso (*Recogida y registro de evidencias evaluativas*) la proliferación de la evaluación continua conlleva la necesidad de articular sistemas y mecanismos para el registro y archivo de evidencias que permitan el seguimiento.

2.1.5.3 Evaluación basada en criterios y explícita.

“Identificar y definir competencias y fijar en su *continuum* el nivel de desarrollo deseable es importante con el fin de comunicar a los estudiantes qué se pretende alcanzar con el proceso de enseñanza-aprendizaje” (Mateo & VLachopoulos, 2013, p. 190). En consecuencia, evaluar competencias exige establecer sistemas de evaluación criterial, que reflejen en cada caso el desempeño alcanzado por las/os estudiantes (Rué, 2008; Sagasta & Bilbatua, 2009). Salvo en aquellos casos en los que se persiga la selección, Biggs (2005) desaconseja el uso de la evaluación normativa, por sus limitaciones en la certificación de resultados. Dada su complejidad, en la evaluación de competencias resulta todavía más necesario el establecimiento de una serie de criterios que reflejen y operativicen los logros a alcanzar (Villa & Poblete, 2007, 2011).

“El alumnado debería conocer cuáles son criterios de calificación, quién será su agente evaluador y los propósitos a los que servirá la evaluación” (Baartman, Bastiaens, Kirschner, & Van der Vleuten, 2006, p. 163). Por ese motivo, el enfoque evaluativo basado en competencias ha promovido la generalización de las rúbricas o matrices de evaluación. Concebidas como herramientas o instrumentos al servicio de docentes y discentes, éstas favorecen la explicitación de los criterios de evaluación contribuyendo a la transparencia general del proceso (Andrade, 2005; Barberá & Martín, 2009; Blanco, 2008b; Coll, Rochera, & Onrubia, 2009; Panadero & Jonsson, 2013) y, en consecuencia, los procesos de retroalimentación (Atkinson & Lim, 2013).

2.1.5.4 Evaluación práctica y contextualizada.

El carácter práctico de las competencias provoca que su aprendizaje y evaluación sólo pueda tener lugar en acción (Cano, 2007; Gairín, 2009; Navío, 2005; Unigarro & Rondón, 2005). En este contexto, los clásicos formatos evaluativos teóricos, basados en la verbalización de saberes, pierden relevancia (Monereo, Castelló, & Gómez, 2009). En su lugar, emerge la inquietud por diseñar tareas evaluativas que propongan situaciones prácticas, con sentido fuera del contexto meramente académico (I. Álvarez & Gómez, 2009;

Baartman et al., 2006, 2007; McDowell & Sambell, 2003; Rué, 2008). En opinión de Cano e Ion (2012), Badía y Álvarez (2009) o Coll et al. (2009), no puede ser de otra manera, ya que el desarrollo de competencias constituye un aprendizaje sistemático y situado.

Dado que la competencia sólo puede evaluarse en acción, una evaluación auténtica debe colocar al alumno en situaciones lo más parecidas posible a la vida real y presentarle problemas y desafíos que deberá afrontar en su actividad profesional, para cuya resolución deberá activar un conjunto integrado de conocimientos, de habilidades y de actitudes y saberlas aplicar estratégicamente en contexto concreto”. (Sayós, 2010, p. 131).

En coherencia con la conceptualización establecida en el capítulo segundo, en la evaluación contextualizada caben dos grandes formatos en función del grado de autenticidad de la situación presentada en la tarea: la evaluación de la ejecución o desempeño y la evaluación auténtica (Durán, 2009; Meyer, 1992b). De creciente proliferación, las TIC también facilitan la simulación de tareas reales para el desarrollo y la evaluación de competencias (Bezanilla et al., 2014; Guenaga et al, 2013). En cualquiera de los formatos, las situaciones o tareas de evaluación contextualizada deben desarrollarse considerando una abundante cantidad de condiciones descritas desde la literatura. La tabla 42 presenta una síntesis de las más relevantes:

Tabla 42
Características principales de la evaluación contextualizada

CONDICIONES PARA EL DISEÑO DE TAREAS O SITUACIONES EVALUATIVAS CONTEXTUALIZADAS	
• Deben plantear demandas similares a las requeridas en el contexto profesional. Eso implica que las situaciones propuestas serán complejas , es decir...	✓ Tareas variadas que exigen la transferencia, no la reproducción.
	✓ Tareas que exigen movilización de recursos de naturaleza diversa.
	✓ Tareas divergentes que no admiten una única alternativa de solución.
	✓ Tareas que exigen juicio y discernimiento.
• Deben apoyarse en criterios de evaluación similares a los que se aplican en la práctica profesional real. Eso lleva a...	
	✓ Tareas fundamentalmente cualitativas, centradas en la calidad del desempeño.

CONDICIONES PARA EL DISEÑO DE TAREAS O SITUACIONES EVALUATIVAS CONTEXTUALIZADAS

- Deben ser relevantes y significativas desde la perspectiva formativa. Eso lleva a...
 - ✓ Tareas justificadas dentro de los objetivos formativos previstos en el programa y la titulación.
 - ✓ Tareas incardinadas en la dinámica del aula, relacionadas con el trabajo efectuado.
 - ✓ Tareas dinámicas, previstas con perspectiva y dilatadas en el tiempo.
- Deben desarrollarse en condiciones similares a las que tienen lugar en la práctica profesional real. Eso lleva a...
 - ✓ Tareas que reproducen con la mayor fidelidad posible los contextos de intervención habituales y las relaciones verticales y horizontales que en el mismo tienen lugar.
 - ✓ Tareas apoyadas, en cuya resolución pueden aplicarse las ayudas y recursos habitualmente disponibles en el contexto profesional real.

Nota: Reelaboración propia a partir de varias contribuciones (Ahumada, 2005; I. Álvarez & Gómez, 2009; I. Álvarez, 2009; Baartman et al., (2007); Badía & Álvarez, 2009; Boud, 2009; Coll et al., 2009; Engel & Bustos, 2009; A. Fernández, 2010; Gulikers et al., 2004; McDowell & Sambell, 2003; Monereo, 2009).

2.1.5.5 *Evaluación diversa y divergente.*

La evaluación de competencias se aleja de los patrones evaluativos uniformes y, fruto de su complejidad natural, se aproxima a la diferencia desde distintas perspectivas. Según ha sido anticipado en condiciones anteriores, esta aproximación adopta dos manifestaciones principalmente:

- La evaluación de competencias debe ser *diversa*, de modo que la recopilación de evidencias tenga lugar en situaciones diferentes y por medio de herramientas distintas. Esa variedad favorece la transferencia del aprendizaje más que la reproducción (Engel & Bustos, 2009; A. Fernández, 2010). En línea con las recomendaciones de Boud y asociados (2010), también contribuye a satisfacer la diversidad de necesidades y expectativas coexistentes en un mismo grupo.
- Además la evaluación de competencias debe tender hacia formatos *divergentes*, en los que sea posible más de una perspectiva de análisis y/o resolución (Ahumada, 2002, 2005; Castelló et al. 2009; Dochy et al., 2002). Las competencias están dotadas de un carácter situacional que las hace versátiles y multifuncionales (Cano, 2007; Unigarro & Rondón, 2005; Villa & Poblete, 2007; Viso, 2010). Su movilización se produce en situaciones complejas, en las que rara vez es posible una única alternativa de solución.

2.1.5.6 Evaluación participativa.

El paradigma formativo asociado al EEES sitúa al alumnado y su aprendizaje en el centro de la formación, lo que exige tomar medidas para favorecer su participación activa. De acuerdo con V. M. López (2009a) los formatos participativos de evaluación representan la forma de evaluación más lógica y coherente con el planteamiento del EEES. Ante la consigna de una mayor autonomía discente, conceden protagonismo a las/os estudiantes y les propician experiencias para iniciarse en la autorregulación del aprendizaje, capacidad requerida social y profesionalmente (Hernández et al., 2006; Hernández, Sales, & Cuesta, 2010).

El nuevo sistema de enseñanza universitaria que quiere implantarse, demanda que la respuesta del alumno en el proceso de aprendizaje y evaluación sea un elemento básico a la hora de avanzar hacia aprendizajes más profundos y de mayor calidad. La evaluación adquiere una nueva dimensión al situar al alumno en el centro del proceso de aprendizaje y al aplicarse un enfoque basado en la adquisición de competencias que conlleva un replanteamiento de la docencia y del diseño de todos los elementos estructurales que la conforman". (Ureña et al., 2009, pp. 105-106).

La necesidad y pertinencia de la intervención discente en la evaluación queda respaldada desde dos líneas de argumentación paralelas: por una parte, responde a las implicaciones derivadas de la valoración de competencias (Villa & Poblete, 2007, 2011) y por la otra, representa una recomendación ampliamente referida desde el enfoque teórico de la evaluación para el aprendizaje. En consecuencia, la delimitación de los agentes de la evaluación, pasa a constituir el último de los criterios a determinar en el diseño de la evaluación. Su presentación se realiza a continuación.

2.1.6 Evaluación compartida o participativa.

La implicación del alumnado en la evaluación de su aprendizaje representa uno de los rasgos evaluativos con mayor presencia entre los referentes consultados. Además de estar vinculada a la evaluación de competencias (tabla 41), la intervención discente constituye una de las bases del enfoque teórico de la evaluación para el aprendizaje (ARG, 2002; Carless, 2007; Stiggins, 2005a, 2005b; Stiggins et al., 2006). Su presencia es

destacada entre los antecedentes que han tratado de sistematizar rasgos deseables para la praxis evaluativa universitaria, y así se ha constatado en nueve de las 10 referencias consideradas (tabla 35). La literatura nacional e internacional también respalda la relevancia de la evaluación participativa, aportando numerosos testimonios que justifican su pertinencia y utilidad pedagógica (firmados, entre otros muchos, por Boud, 1995; Boud & Falchikov, 2007; Dochy et al., 2002; Falchikov, 2005; Falchikov & Goldfinch, 2000; V. M. López, 2011; V. M. López et al., 2007; A. Pérez et al., 2008; Shepard, 2000).

Aunque caben modalidades y niveles de implicación muy diversos, las fórmulas participativas hacen de la evaluación un proceso dialógico y/o de toma de decisiones compartidas (J. Sánchez, 2011). A. Brew (2003) diferencia entre aquellas que tienen lugar de modo informal o espontáneo y las formales, deliberadamente estructuradas para provocar la participación discente. El modelo ITC se centra exclusivamente en estas últimas, que hacen de la evaluación una actividad premeditada de aprendizaje diseñada por el profesorado.

De acuerdo con la conceptualización propuesta en el capítulo segundo (**ANEXO 03**), la participación discente puede tener lugar en cualquiera de las fases del proceso de evaluación, en modalidades organizativas diferentes y con o sin efecto en la calificación:

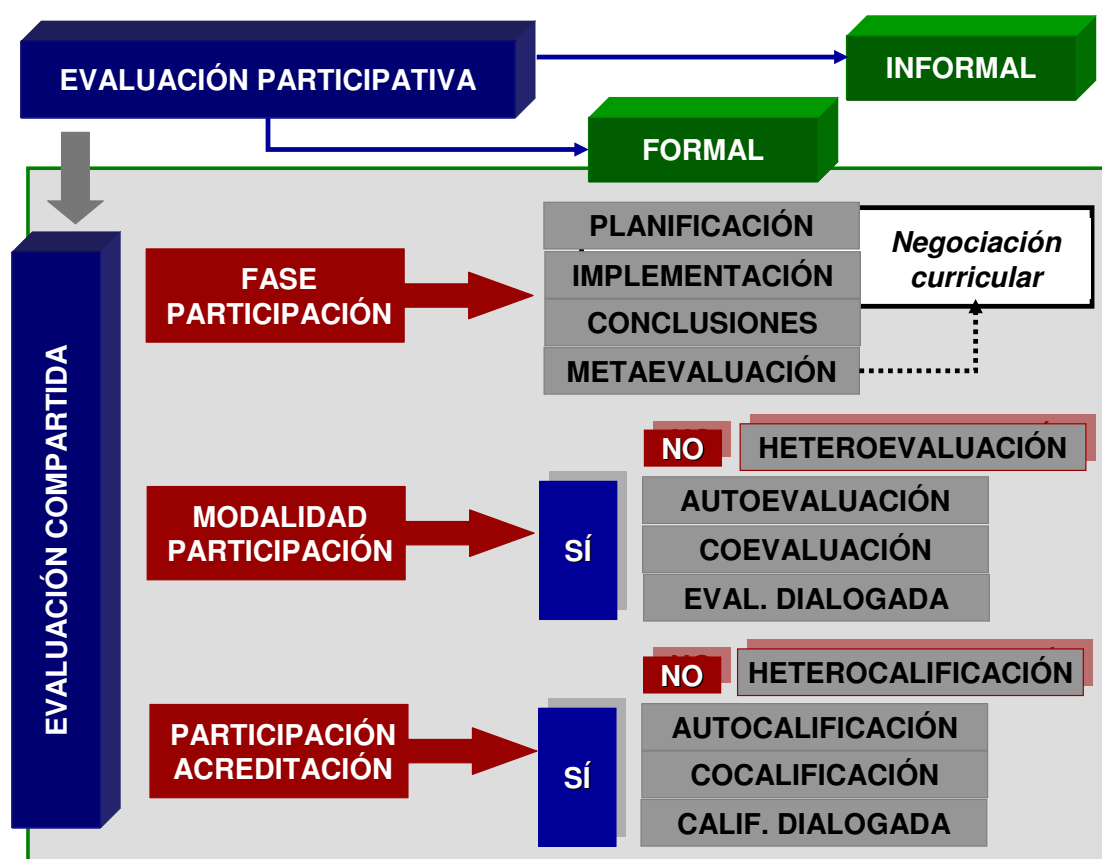


Figura 19. Modalidades de evaluación participativa contempladas en el modelo ITC.

Sintetizadas en la figura 19, las distintas modalidades de evaluación participativa reportan beneficios potenciales relevantes al proceso formativo. Resumiendo los más reseñados, advertimos que este tipo de evaluación conlleva siete ventajas principales:

- Beneficia el desarrollo de la capacidad de autoevaluación y autorregulación del aprendizaje y, en consecuencia, representa una estrategia para promover el aprendizaje autónomo a lo largo de la vida (Boud & Falchikov, 2006, 2007; Buscà, Capllonch et al., 2010; Hernández et al., 2006, 2009; Dochy et al., 1999; O. Esteve, 2010; Ljungman & Silen, 2008; V. M. López, 2006, 2008, 2009). Por ese motivo, resulta particularmente importante en la conceptualización de la evaluación sostenible u orientada al aprendizaje, presentada en el capítulo segundo. “La capacidad para aprender está íntimamente relacionada con la capacidad para autorregular el aprendizaje. Desde que nacemos, las personas vamos construyendo nuestro propio estilo de aprender y lo vamos mejorando con el tiempo” (Sanmartí, 2007, p. 54). En este sentido, “la implicación del alumnado en los procesos de evaluación de su aprendizaje debería ser una herramienta básica de trabajo cuando se pretende desarrollar la autonomía del alumnado y la capacidad de gestionar su propio aprendizaje” (V. M. López, 2012, p. 119).
- Contribuye al desarrollo de la capacidad evaluativa, recurso interdisciplinar y polivalente, aplicable a cualquier campo profesional (I. Álvarez, 2008; Boud, 1995, 2000; Boud et al., 1999; Capllonch et al., 2009; Sluijmans et al., 1999). “Aprender a presentar valoraciones al trabajo y no a la persona y aprender a recibir apreciaciones y críticas constructivas de los demás como elementos para la mejora [...]” (Villardón, 2006, p. 67), supone un aprendizaje relevante que toda/o estudiante universitario debería construir en su formación profesional básica. En este sentido, las fórmulas de evaluación participativa contribuyen a desarrollar capacidades de negociación (A. Brew, 2003), de realización de juicios independientes (Brown & Knight, 1994; Vu & Dall’Alba, 2007) y de toma de decisiones basada en estándares (Sluijmans et al., 1999). Según Ureña et al. (2009) también pueden contribuir al fortalecimiento de los procesos de trabajo en equipo cuando exceden la autoevaluación o autocalificación.
- Favorece la mejor comprensión e interiorización de los criterios de calidad y las condiciones evaluativas, abordada en el próximo subproceso (*Comunicación y difusión de criterios y condiciones*). En la literatura está ampliamente generalizada la idea de que las representaciones que conforma el alumnado sobre los objetivos

de aprendizaje y los criterios para su valoración se ajustan a raíz de su participación en la evaluación (Biggs, 2005; Boud & asociados, 2010; Jordan, 2003; V. M. López 2007, 2009; J. Rodríguez, 2008; Sagasta & Bilbatua, 2009; Sanmartí, 2007, 2010).

- Refuerza la perspectiva didáctica de la experiencia evaluativa y disminuye la sensación de amenaza (I. Álvarez, 2008a; Lu & Law, 2012). En consecuencia, ayuda a equilibrar el impacto emocional de la evaluación sobre el alumnado, otro de los elementos destacados desde el enfoque teórico de la evaluación para el aprendizaje (Arter, 2009; Boud & asociados, 2010; J. Chappuis, 2009).
- Favorece la propia viabilidad o sostenibilidad del sistema de evaluación. Su contribución es particularmente reseñable para aumentar la frecuencia y puntualidad del feedback (Cano, 2014; Carless, 2013a; Carless et al., 2011; Liu & Carless, 2006). Así lo evidenciaron C. Brew et al. (2009) al determinar que, mientras que las/os estudiantes aducen motivos pedagógicos para defender la importancia de su implicación en la evaluación, una parte importante del profesorado apela a cuestiones prácticas como el ahorro de tiempo.
- Contribuye al incremento de los niveles de motivación y los resultados de aprendizaje (Dochy et al., 1999; Falchikov & Boud, 1989; Ibarra et al., 2012; Joughin, 2003; V. M. López et al., 2007; Ureña et al., 2009). Efectuado un importante trabajo de revisión sobre el tema, Bretones (2008) concluyó que el uso de evaluación participativa mejora la satisfacción del alumnado y también su aprendizaje. En opinión de I. Álvarez (2008a), esta mejora se debe a que la experiencia de evaluación favorece una imagen ajustada del objetivo a alcanzar y las condiciones para su valoración.
- De forma indirecta, da lugar a procesos de reflexión que pueden conducir al *desarrollo profesional del profesorado participante* (V. M. López et al., 2007; Ramsden, 2003). Ureña et al. (2009) consideran que “cuando los profesores comparten con sus estudiantes el proceso de evaluación, delegando control, compartiendo poder y fomentando que los alumnos puedan evaluarse a sí mismos, se realza el juicio profesional de ambos” (p. 110).

Los beneficios reseñados son aplicables a cualquier titulación universitaria, pero adoptan una especial significación en los grados dedicados a la formación básica de maestras/os de Educación Infantil y Primaria. Este tipo de experiencias permiten al alumnado situarse en el rol docente que habrá de ejercer en el futuro y comprender la

naturaleza emocional de la evaluación (I. Álvarez, 2008; C. Brew et al., 2009; V. M. López, Castejón, & Pérez, 2012; McGarr & Clifford, 2013).

El hecho de dialogar y compartir con el alumnado el proceso de evaluación y calificación, incluidos los mecanismos de control y poder que encierra la calificación, fomenta considerablemente los procesos metacognitivos sobre el aprendizaje, pero, sobre todo, ayuda a ir cambiando la perspectiva y rol adoptado por el alumnado, que comienza a entender los procesos de aprendizaje como responsabilidad e interés propio de cara a su futura profesión, en vez de como unos simples conocimientos que ha de memorizar de cara a un examen. Nos solemos referir a este proceso como la evolución del rol de alumno al rol de profesional en formación inicial. (V. M. López et al., 2007, p. 12).

Pero el desarrollo de fórmulas participativas de evaluación también está sujeto a una serie de **dificultades** que se le asocian de forma habitual. La revisión efectuada sobre el tema señala obstáculos relacionados con la inexperiencia docente y discente y con sus respectivas tensiones y resistencias. El diseño de dispositivos evaluativos debe partir de la consideración expresa de los más habituales para neutralizar, en lo posible, los efectos indeseados:

- Como indica Margalef (2009) la autoevaluación y la coevaluación no forman parte de la cultura evaluativa del alumnado, lo que conlleva una capacidad evaluativa insuficientemente cultivada.
- La implicación del alumnado en la evaluación está rodeada de dudas y recelos sobre su rigor y fiabilidad. En la literatura específica los argumentos y testimonios en favor y en contra se suceden e intercalan (Bretones, 2008). Existe un importante número de estudios que revelan niveles razonables de fiabilidad en los procesos de evaluación o calificación compartida (Dochy et al., 1999; Falchikov & Boud, 1989; Flemming, 2003; V. M. López, Castejón, et al., 2012; V. M. López, Fernández-Balboa, & Santos, 2012; Pope, 2005; Race, 2003; Vu & Dall'Alba, 2007; Watts & García-Carbonell, 2006). Junto a ellos, también son frecuentes las referencias a las amenazas al rigor y la objetividad en valoraciones producidas por estudiantes (Capllonch et al., 2009; Roach, 2003; Ureña et al., 2009). Aunque estas desviaciones tienden a atribuirse más a la inexperiencia que a prácticas

deshonestas (Lapaham & Webster, 2003; G. Rodríguez et al. 2012), su incidencia genera recelos y resistencias.

- Las prácticas evaluativas participativas son costosas para el profesorado por el tiempo y esfuerzo que exige su desarrollo y por el desgaste emocional que produce la inexperiencia en las mismas (dificultad para anticipar necesidades u obstáculos, nuevo rol, etc.) (Margalef, 2009; Zaragoza et al., 2009).
- Los formatos de evaluación que requieren la participación del alumnado y, especialmente, los que tienen lugar de forma grupal, pueden acarrear fricciones y tensiones (Heathfield, 2003; McGarr & Clifford, 2013; Vu & Dall'Alba, 2007) generando una fuente de estrés considerable (Pope, 2005). Así lo constató McMahon (2010) en su estudio, detectando entre el alumnado resistencias y desinterés por participar en procesos de evaluación participativa. Tener control o potestad sobre el trabajo de otras/os compañeras/os representa un hecho impopular (Falchikov & Boud, 1989). Las/os estudiantes no consideran que la evaluación sea parte de su tarea, no se sienten preparadas/os para calificar el trabajo ajeno y declaran incomodidad al respecto (Boud et al., 1999; Falchikov, 2005; Hamodi & López, 2012; Ibarra et al., 2012; Wilson, Diao, Huang, 2015). Por todo ello, es cada vez más frecuente la recomendación de sesiones de formación o entrenamiento previo, que mitiguen las resistencias y dificultades iniciales (A. Brew, 2003; Jordan, 2003; Lu & Law, 2012; Vu & Dall'Alba, 2007).
- El desarrollo de prácticas evaluativas participativas también *suscita recelos y tensiones en el colectivo docente*, en particular, cuando éstas inciden en la calificación. El debate de fondo se produce al contemplar los riesgos que pudiera ocasionar el mal uso de las modalidades de calificación compartida, frente a la responsabilidad profesional de garantizar que toda/o graduada/o esté plenamente capacitada/o para ejercer la profesión (Capllonch et al., 2009).

2.2 Comunicación y difusión de criterios y condiciones.

El subproceso segundo atiende a la presentación de los criterios y condiciones en los que se basa la evaluación, así como a las medidas que se adoptan para su difusión y profundización entre el alumnado. En su desarrollo incorpora dos grandes criterios o requisitos interrelacionados, que suscitan un amplio acuerdo entre los referentes

consultados, aunque con desigual énfasis. El primero de ellos plantea un nivel básico de información o explicitación, mientras que el segundo se orienta a la estimulación deliberada de su interiorización por parte de los agentes evaluados.

2.2.1 Explicitación y difusión de criterios y condiciones.

Afirma Sanmartí (2007) que,

la idea que los alumnos tienen de lo que han de aprender no depende tanto de lo que el profesorado les dice, sino de lo que éste tiene realmente en cuenta en el momento de evaluar, y con relación a ello adaptan su forma de aprender. (p. 22).

En consecuencia, detallar y explicitar cuáles son los criterios en los que se apoya la evaluación representa una importante herramienta didáctica para enfocar deliberadamente el esfuerzo y el trabajo discente.

Como resume Stobart (2010), “el fundamento teórico es claro: es más fácil aprender cuando sabemos lo que hacemos y adónde estamos intentando llegar” (p. 181). Así se reconoce desde el enfoque teórico de la evaluación para el aprendizaje (ARG, 2002; Carless, 2007; J. Chappuis et al., 2012; Stiggins, 2005a, 2005b) y en los diez decálogos o manifiestos para la calidad evaluativa analizados (tabla 35). Directamente vinculado al derecho a la información que asiste al alumnado, este requisito se encuentra especialmente representado entre los referentes normativos (ENQA, 2009; Real Decreto 1791/2010), en las propuestas estandarizadas (Baker et al., 1996; CRESST, 2012; JCSEE, 2007; Land, 1997) y en los sistemas institucionales para la calidad evaluativa (Astin et al., 1992; QAA, 2011, 2014).

Introducida por N. Frederiksen (1984), el énfasis en la transparencia evaluativa adopta matices específicos en el contexto evaluativo actual.

- La complejidad propia de las competencias como objeto de evaluación hace especialmente necesario que se detallen con claridad los logros previstos, estableciendo distintos niveles de consecución (Villa & Poblete, 2007, 2011). Su desarrollo es progresivo, de modo que las/os estudiantes precisan de referencias que les ayuden a comprender y prever los distintos estadios en su evolución:

El hecho de que el aprendizaje de las competencias se pueda ir perfeccionando, conlleva la necesidad de evaluar el grado de eficiencia con

la que las competencias se aplican. Por ello las pruebas de evaluación deberán incorporar indicadores de logro que permitan evaluar el grado de eficacia con la que se resuelven las situaciones planteadas. (Arnau, 2009, p. 1).

- Desde la perspectiva teórica de la evaluación para el aprendizaje, Nicol y Macfarlane (2006) y Gibbs y Simpson (2009) enfatizan la importancia de que además de claras, las expectativas comunicadas sean elevadas, y así se plasma en el Proyecto REAP (2005). La evaluación debe fomentar creencias motivacionales positivas para movilizar el esfuerzo y el trabajo del alumnado en dirección a los logros establecidos (Nicol, 2007, 2009). Desde esta convicción, Biggs (2005) y Sans (2008) relacionan exigencias evaluativas de mayor nivel con producciones discentes de mayor calidad.
- Dentro del marco general de la evaluación participativa, los procesos de negociación curricular pueden representar una vía efectiva para clarificar las expectativas evaluativas, de modo que los efectos de repercusión (Biggs, 2005) puedan tener lugar en la dirección a ellas. Bretones (2008), por ejemplo, concluyó que el rigor de las valoraciones efectuadas por estudiantes mejora cuando los criterios de evaluación han sido previamente negociados⁵¹, apreciación general que también comparten Capllonch y Buscà (2012).

2.2.2 Comprensión e interiorización de las condiciones básicas.

La explicitación de criterios y condiciones constituye un requisito indispensable para la delimitación de expectativas ajustadas, pero es insuficiente por sí mismo. Para desarrollar todo su potencial, la evaluación no sólo debe ser transparente, sino que además debe procurar activamente que las/os estudiantes comprendan e interioricen aquello que se espera de su trabajo (J. Chappuis, 2005; Price & O'Donovan, 2006; Sadler, 1989).

Las representaciones que se forme el alumnado sobre las exigencias evaluativas ejercen una influencia determinante en cómo se dirige y desarrolla su esfuerzo y su trabajo

⁵¹ Como ya ha sido justificado, para preservar la especial relevancia de la evaluación criterial en el contexto de las profesiones reguladas, el modelo ITC acota la posibilidad de intervención discente en el diseño evaluativo a la definición de las condiciones. Aunque no se contemple la negociación de los criterios de evaluación, consideramos que el trabajo que genera la evaluación participativa sobre los mismos puede contribuir de forma eficaz a su clarificación y comprensión.

(Biggs, 2005; MacLellan, 2001; Meyer & Watson, 1991; Ramsden, 2003), de modo que es importante procurar que sean correctas y ajustadas. Éstas se conforman en todo caso, a partir de la observación más o menos consciente de las situaciones habituales de trabajo (Castillo, 2004; Gibbs, 2003).

Independientemente de lo que esté claramente especificado en la documentación de la asignatura que se pretende evaluar, los estudiantes averiguan por sí mismos qué es lo que cuenta-o, por lo menos, lo que ellos creen que cuenta- y dirigen sus esfuerzos de manera acorde. (Gibbs & Simpson, 2009, p. 10).

Pero a pesar de los intentos del profesorado por transmitir con claridad las expectativas de aprendizaje y las condiciones para su valoración, hay estudiantes que tienen serias dificultades para comprenderlas (Carless, 2002, 2006; Sadler, 2010, 2013).

El profesorado puede intentar comunicar los objetivos explicitándolos verbalmente o por escrito, insistir en la necesidad de pensar-planificar antes de actuar y animar a aplicar criterios de evaluación antes de dar por terminada una actividad, pero a los alumnos les es difícil reconocer la importancia de todo ello y, muy especialmente, de que sea necesario para su aprendizaje. (Sanmartí, 2010, p. 28).

La falta de herramientas personales para decodificar los estándares de logro propios de la disciplina (Higgins et al., 2002; Sambell, 2011; Price et al., 2010), la interferencia de las representaciones iniciales (A. Calatayud, 2008; Otero, 2000) o un simple desajuste en la interpretación de las demandas (Zaragoza et al., 2009) pueden ser causa frecuente de esa incompreensión. En consecuencia, el modelo ITC prevé la realización de acciones específicas para favorecer la comprensión e interiorización de los criterios de evaluación.

Durante el proceso formativo se debe dotar al alumnado de los apoyos e instrumentos que le capaciten para entender el significado de los objetivos y evaluar su situación de partida y su progreso respecto a los mismos (O. Esteve, 2010). Para ello es posible recurrir a alternativas diversas y complementarias; junto a las rúbricas o la participación directa del alumnado en la evaluación, la comisión de feedback representa otra estrategia que contribuye al ajuste de interpretaciones (I. Álvarez & Olivera-Smith, 2014; Atkin et al., 2001; Harrington et al., 2006; Jessop, 2013; Laurillard, 2002; Ramaprasad, 1983).

2.3 Recogida y registro de evidencias evaluativas.

El modelo ITC detalla un tercer subproceso de recopilación y registro de evidencias dentro del proceso evaluativo. Ubicado en la fase de implementación o puesta en práctica, su desarrollo debe tener lugar de acuerdo a dos requisitos según los referentes consultados.

2.3.1 Implementación ajustada.

La adopción de decisiones conscientes y justificadas en la fase de diseño debe dar lugar a una implementación igualmente consciente del sistema definido, que no se desvíe de lo previsto, a menos que sea de forma justificada. Por un lado, dispositivos evaluativos estratégicamente ideados para promover aprendizaje podrían resultar ineficaces al desviarse total o parcialmente las pautas y condiciones determinadas para ello. Por otro lado, el incumplimiento de la previsión también podría llegar a amenazar la transparencia y publicidad del proceso evaluativo, por contravenir la información explicitada. De hecho, los referentes normativos para la evaluación del aprendizaje en la enseñanza universitaria muestran el cumplimiento coherente de los compromisos evaluativos publicados como responsabilidad docente y derecho del alumnado (ANECA, AQU, & ACSUC, 2007b; ENQA, 2009; Real Decreto 1791/2010). Esta inquietud se encuentra especialmente extendida en las universidades del Reino Unido por impulso de las directrices de la QAA (2014).

Sin embargo, un diseño evaluativo cuidadosamente delimitado también puede quedar desvirtuado si no se ofrece respuesta efectiva a los imprevistos advertidos en su puesta en práctica.

Los procesos evaluativos en educación nunca deben definirse sobre la base de lineamientos rígidos y dogmáticos, en realidad un plan de evaluación educativa debe prever desde el inicio márgenes de flexibilidad en su desarrollo a sabiendas de que difícilmente todo se producirá como se había previsto. El mejor evaluador será, pues, aquél que sea capaz de proponer cambios y alternativas frente a los desajustes y sucesos que puedan ocurrir. (Monereo & Castelló, 2009a, p. 82).

Así lo considera también J. Rodríguez (2008), quien defiende la realización de planificaciones evaluativas flexibles, que dejen espacio y cabida a los ajustes y adaptaciones que deban tener lugar durante el proceso de desarrollo.

En balance de ambas consideraciones, el modelo ITC propone la puesta en práctica coherente del dispositivo evaluativo diseñado y la gestión adecuada de los cambios que se realicen en el transcurso de la misma. Ello exige la comunicación de las adecuaciones introducidas, de modo que el alumnado siga estando al tanto de los criterios y procedimientos en función de los cuales se va a valorar su trabajo. También la justificación de dichas modificaciones, de forma que se haga partícipes a las/os estudiantes del proceso evaluativo y se les introduzca paulatinamente en el juicio profesional que deberán manejar en un futuro próximo (I. Álvarez, 2008; V. M. López et al., 2007; Ureña et al., 2009). De acuerdo con Shepard (2000) compartir con el alumnado las investigaciones y deliberaciones que el profesorado plantee sobre su enseñanza representa una experiencia formativa. Tanto más en la formación básica de maestras/os, donde el alumnado actual deberá experimentar en primera persona las dificultades de la implementación de los dispositivos evaluativos.

2.3.2 Recopilación y registro personalizado de evidencias de aprendizaje.

El impulso del enfoque teórico de la evaluación para el aprendizaje y la apuesta del EEES por el desarrollo de competencias han favorecido que la evaluación deje de concebirse como un hecho puntual, ubicado en el tramo final del proceso formativo. La proliferación de dispositivos de evaluación continua pone de manifiesto, más que nunca, la necesidad de hacer un seguimiento del grupo de estudiantes y sus integrantes, lo que exige el desarrollo de sistemas efectivos de registro y archivo de información y evidencias (Coll et al., 2007; Escamilla, 2010). De acuerdo con el enfoque de trabajo interdisciplinar y globalizado al que da lugar el desarrollo de competencias, este registro debería abarcar idealmente toda la trayectoria formativa:

En un currículum donde las habilidades se demuestran a través de una variedad de unidades académicas, y donde los estudiantes pueden optar por una amplia variedad de ellas, es importante que un sistema recoja juntos todos los aspectos del progreso individual y del logro a lo largo del programa. Particularmente útil es la confección de un perfil, que no es un método de evaluación, sino un modo

multidimensional de presentar una información sobre los logros de un estudiante y su actuación en una amplia variedad de conocimientos, habilidades y cualidades personales. (Rhodes & Tallantyre, 2003, p. 137).

Dicha aspiración se materializa en el Proyecto HEAR (Burgess, 2012), que ofrece a las universidades del Reino Unido un recurso electrónico para reportar los logros que va desarrollando el alumnado a lo largo de su formación superior. A pesar de compartir la necesidad de realizar la observación y registro longitudinal del aprendizaje de cada estudiante, Capllonch et al. (2009) advierten de los riesgos de burocratización y sobrecarga docente asociados.

Para toda la titulación o circunscritos a un determinado módulo o asignatura, los mecanismos articulados para el seguimiento y orientación personalizada deben preservar los derechos legales que asisten al alumnado en cuanto a la protección de su identidad (GRAPA-RIMA 2012; NUS 2010; Real Decreto 1791/2010). Las propuestas estandarizadas para la evaluación del aprendizaje discente se muestran particularmente sensibles a esta necesidad (Baker et al., 1992; JCSEE, 2007; Land, 1997).

2.4 Análisis de resultados.

El cuarto subproceso delimitado en el modelo ITC se centra en el análisis de las evidencias recopiladas para la valoración pedagógica o acreditativa del aprendizaje discente. Ubicado en la fase de puesta en práctica o implementación, el análisis debe atender a dos criterios principalmente: el primero de ellos plantea la adopción de medidas para proteger del efecto de los sesgos en general, mientras que el segundo atiende específicamente al proceso de calificación.

2.4.1 Control de sesgos.

La definición establecida en el capítulo segundo, reconoce la evaluación como forma de investigación y, como tal, no puede renunciar a los supuestos generales de validez y fiabilidad (Albert, 2009). La inquietud por la aceptabilidad técnica de la evaluación representa una constante en los distintos referentes que han sido consultados para la

definición del modelo ITC. Sin negar la complejidad natural del proceso evaluativo y la conveniencia de profundizar en su vertiente pedagógica, existe un reconocimiento unánime de la necesidad de garantizar evaluaciones válidas y fiables del aprendizaje. Ambos atributos representan una referencia recurrente en los decálogos publicados sobre evaluación (Brown et al., 1996; Boud & asociados, 2010; Buscà, Capllonch et al., 2010; Cano et al., 2010; Suskie, 2006) y están ampliamente recogidos en las propuestas que delimitan estándares para su puesta en práctica (Baker et al., 1996; JCSEE, 2007; Land, 1997).

La *validez* es un concepto complejo en el que se discriminan hasta cinco modalidades diferentes (American Educational Research Association- AERA, American Psychological Association- APA, & National Council on Measurement in Education- NCME, 1999). Sin embargo, los referentes consultados, devuelven una visión general del atributo, entendido como capacidad de la evaluación para valorar aquello que pretende.

Para poder servir a sus propósitos acreditativos o pedagógicos, resulta fundamental que la evaluación valore realmente aquello que pretende comprender, y no cuestiones menores relacionadas (Leal, 2004; McMillan, 2000; Race, 2003). En este sentido, la evaluación válida es aquella que verdaderamente revela en qué medida se han alcanzado los objetivos de aprendizaje previamente establecidos (L. Prieto, 2007).

Situado en el contexto específico de la evaluación de competencias, De la Orden (2011b) ve en la validez un requisito tan fundamental, como complejo de garantizar.

Evidentemente, la condición fundamental de un sistema de evaluación es que sea educativamente válido, lo cual implica no solamente que los instrumentos de medida utilizados (test, pruebas, escalas de observación, procedimientos no formales, etc.) sean válidos en el sentido tradicional, sino también que exista un máximo de congruencia entre los criterios y modos de evaluación, la especificación de objetivos y el proceso didáctico o de intervención educativa. (p. 4).

Mateo (2000) sostiene una postura similar, argumentada desde la complejidad connatural de la evaluación y su duplicidad de objetivos pedagógicos y acreditativos. “Siempre habrá que mantener presente la consecución de ambos objetivos, una evaluación incapaz de asignar adecuadamente las valoraciones correspondientes no gozará de validez, una evaluación de la que no se deriven acciones de mejora perderá su credibilidad” (p. 23).

Algo similar sucede con la *fiabilidad*. A pesar de la variedad de perspectivas que suscita su interpretación, en los referentes consultados el término responde a la idea general del grado de precisión en la medida (Morales, 2007). En el mismo sentido, V. M. López (2006, 2009a) formula la veracidad como principio rector de la evaluación, indicando que ésta debe ser rigurosa, contrastable y basada en evidencias. Sin embargo, “cuanto más nos alejemos de la evaluación de unos rendimientos sencillos, como el recuerdo de informaciones, menos factible resulta hacer juicios fiables” (Knight, 2005, p. 187). Así se constata en los últimos años, al adoptar las competencias como foco central de la formación y la evaluación.

Sin perjuicio de las reservas presentadas, el modelo ITC establece que los dispositivos evaluativos de calidad integran mecanismos y medidas para favorecer conclusiones razonablemente válidas y fiables en las cuatro fases de su desarrollo:

- La consolidación de prácticas válidas y fiables depende, en primer lugar, de decisiones relacionadas con la alineación de la evaluación y su instrumentación, adoptadas en la fase de planificación. Exige una correcta formulación de indicadores y criterios de evaluación⁵², una selección adecuada de las evidencias que pueden manifestar su logro y la previsión de herramientas y tareas capaces de captar con precisión diversos grados de consecución.
- La puesta en práctica del dispositivo evaluativo es la menos previsible de las fases y, en consecuencia, la más vulnerable a las amenazas que pueden invalidarlo o restarle credibilidad. Desde la previsión y la actuación consciente, deben articularse medios para reducir el riesgo de situaciones como las expuestas a continuación a modo de ejemplo:
 - ✓ Las condiciones en las que tiene lugar la recopilación de evidencias pueden favorecer prácticas evaluativas más o menos válidas; la forma de plantear tareas, el orden de intervención de las/os diferentes estudiantes u otros factores circunstanciales incontrolados pueden provocar que lo que capte la evaluación no sea precisamente el aprendizaje pretendido.
 - ✓ Los plagios o demás formas de fraude, representan otra amenaza evidente a la fiabilidad del proceso evaluativo (Boud & asociados, 2010; CLE-UB, 2011; CSHE, 2002; Sabirón & Arraiz, 2013).

⁵² Esta corrección hace referencia tanto a la perspectiva formal de la claridad, como a la de la coherencia o alineación con los propósitos del módulo/assignatura y los de la titulación.

- El análisis de las evidencias recopiladas también puede incurrir en desviaciones que disminuyan el ajuste de la evaluación a los objetos previstos y su precisión, de modo que deben preverse y neutralizarse en lo posible. Factores como el orden de corrección de los trabajos, su presentación o las expectativas docentes sobre el/la autor/ora⁵³ puede dar lugar a sesgos de imprecisión (Biggs, 2005; Flemming, 2003; Santos, 1999, 2010). Leal (2004), por su parte, advierte de la fatiga de quien evalúa en los procesos de corrección y de los efectos que puede ocasionar en las valoraciones su estado de ánimo (desviaciones de los criterios de partida, excesiva exigencia o benevolencia, etc.).
- La revisión del dispositivo también debe aportar evidencias para valorar sus posibilidades y limitaciones técnicas y contribuir a mejorar la adecuación y precisión. Así se contempla en el subproceso séptimo (*Revisión del proceso de evaluación: metaevaluación*).

2.4.2 Proceso de calificación.

Sayós (2010) afirma que el énfasis sobre la evaluación continuada y pedagógica no descarta su función acreditativa, necesaria para responder a la responsabilidad social de comprobar la eficacia del proceso de enseñanza-aprendizaje y la calidad de la institución. Los atributos presentados en el anterior criterio son necesarios en toda evaluación que pretenda desarrollarse, pero adoptan una especial relevancia cuando se prevé el uso acreditativo de las conclusiones (Gibbs, 2003) o de alto riesgo (Knight, 2005). Desde esta perspectiva, el modelo ITC establece que los dispositivos evaluativos de calidad deben procurar procesos de calificación que transcurran sin vulnerar los compromisos normativos y los derechos de las/os agentes involucrados. Su desarrollo debe tener lugar en condiciones metodológicas que aseguren su adecuación y fiabilidad, dentro de unos límites razonables. Incorpora para ello tres criterios referidos a otros tantos atributos deseables: la coherencia, la racionalidad y la transparencia y fiabilidad.

En toda evaluación se deben propiciar las condiciones que permitan distinguir entre el juicio arbitrario y el juicio razonable (Goñi, 2005). Así se refleja en los referentes normativos consultados, en los que dicha racionalidad aparece ligada a la objetividad y la transparencia (ANECA, AQU, & ACSUC, 2007b; ENQA, 2009; Real Decreto 1791/2010). Su análisis evidencia que los procesos de calificación deben regirse de acuerdo a criterios

⁵³ Efecto halo, condicionado a su vez por otras variables sociodemográficas como el género o la etnia.

públicamente explicitados y en coherencia con los compromisos adoptados en el propio sistema de evaluación. Las propuestas estandarizadas consideradas también se muestran muy sensibles con estas garantías; la propuesta del JCSEE (2007) las contempla dentro de los estándares de adecuación y precisión y las presentadas desde el CRESST (Baker et al., 1996; CRESST, 2012; Land, 1997) entre los de justicia y credibilidad.

Estrechamente unido a lo anterior, la precisión o fiabilidad de los procesos de calificación del aprendizaje representa una constante en la literatura evaluativa, pero ha cobrado una especial relevancia con el planteamiento del EEES. Como señalan Valcárcel (2003) o Villardón (2006), uno de los grandes retos vinculados al mismo es la promoción de la movilidad de las/os estudiantes, lo que exige el desarrollo de sistemas de evaluación transparentes y rigurosos, que favorezcan el reconocimiento del aprendizaje efectuado y faciliten procesos adecuados de integración y adaptación⁵⁴. Esta necesidad reaviva una duda y un debate tradicional con más fuerza que nunca, por la incorporación de las competencias como objeto de evaluación. La definición e interpretación clásica de la fiabilidad se sitúa muy próxima a la objetividad, pero la evaluación del aprendizaje de las mismas no se ajusta a la lógica objetivista. Con todo, la calificación no puede instalarse en la imprecisión porque comporta repercusiones importantes para el alumnado, la comunidad profesional y la sociedad. Ante este dilema, el modelo ITC plantea una lectura intermedia de estas condiciones y determina que la calificación precisa es aquella que discrimina distintos niveles consecución en los aprendizajes previstos, según el logro evidenciado en los criterios de evaluación y calificación establecidos en la fase de diseño.

2.5 Feedback y proacción.

El quinto subproceso evaluativo centra su atención en la comunicación de los resultados de aprendizaje y en el feedback y proacción que tiene lugar a partir de la misma.

⁵⁴ Uno de los elementos que mejor manifiesta esta necesidad es el SET o Suplemento Europeo al Título, documento académico llamado a detallar con el máximo rigor posible las competencias desarrolladas por cada estudiante a lo largo de su formación universitaria. Según establece el artículo 3 del Real Decreto 1044/2003 que lo regula, es el documento que acompaña a cada uno de los títulos universitarios de carácter oficial y validez en todo el territorio nacional, con la información unificada, personalizada para cada titulado universitario, sobre los estudios cursados, los resultados obtenidos, las capacidades profesionales adquiridas y el nivel de su titulación en el sistema nacional de educación superior (Art, 3).

Incorpora tres aspectos o actuaciones evaluativas clave entre las que se produce una relación de interdependencia:

- El subproceso arranca con la *comunicación de las conclusiones* a las que haya dado lugar la evaluación, independientemente de que tenga efecto pedagógico o acreditativo. Los siguientes niveles parten de esta información inicial y plantean niveles progresivos de análisis, comprensión y profundización.
- El primero de ellos es el *feedback o retroalimentación* y tiene que ver con la comprensión verdadera del nivel de calidad alcanzado en intentos o actuaciones realizadas. Si bien no existe una definición generalmente asumida para el término (Evans, 2013), entendemos por feedback toda forma de apoyo y ayuda que recibe la/el estudiante tras una demostración total o parcial de su aprendizaje, para comprender en qué lugar se encuentra con respecto a los estándares de calidad o el nivel de referencia y así poder mejorar. De modo similar, Boud y Molloy (2013) lo definen como el proceso por medio del cual “los aprendices obtienen información sobre su trabajo, para apreciar las diferencias y similitudes entre los estándares de idoneidad establecidos para una tarea dada y las cualidades del propio trabajo, y producir así un trabajo mejorado” (p. 6).
- Sólo si se comprende la situación o estado actual respecto al aprendizaje a realizar es posible avanzar al segundo nivel, la **proacción**. Asociada a las decisiones de mejora o pautas de regulación desprendidas de la evaluación, ésta puede adoptar dos focos de atención complementarios, recogidos en dos criterios del modelo ITC:
 - ✓ *El aprendizaje o la actuación discente*. Las conclusiones reveladas pueden dar lugar a orientaciones sobre qué debe hacer y cómo debe actuar el alumnado para mejorar su logro en próximos intentos. Se genera por medio del feedback, aunque no todo feedback incluye pautas para la proacción o mejora del aprendizaje (Nicol, 2010, 2013).
 - ✓ *La enseñanza o la actuación docente*. Las conclusiones reveladas también pueden dar lugar a orientaciones sobre la actuación docente, de modo que sugieran mejoras o ajustes en la enseñanza para contribuir a un mejor aprendizaje.

2.5.1 Comunicación de los resultados de aprendizaje.

La comunicación de los resultados de aprendizaje representa el primer nivel de actuación que tiene lugar finalizada una evaluación global o parcial del aprendizaje. Se considera insuficiente por sí misma para contribuir a la mejora, pero debe efectuarse en condiciones legales que respeten las directrices emanadas de la normativa institucional (ANECA, AQU, & ACSUC, 2007b; QAA, 2011), así como los derechos que al respecto asisten al alumnado, principal audiencia interesada (Baker et al., 1996; GRAPA-RIMA 2012; JCSEE, 2007; Land, 1997; NUS, 2010; Real Decreto 1791/2010).

2.5.2 Retroinformación o feedback.

Hace casi dos décadas Hattie, Biggs y Purdie (1996) desarrollaron un estudio sobre la efectividad de las intervenciones docentes directas en el aprendizaje, concluyendo que el feedback o la retroalimentación era la más relevante de todas. Transcurridas casi dos décadas, sigue siendo un núcleo temático ampliamente representado en la literatura, lo que denota su importancia y el interés que suscita a diversos niveles (Carless, 2013; Jessop, 2013; Nicol, 2010; Nicol, Breslin, & Thompson, 2014; OGG-ASKe, 2009; Orsmond, Maw, Park, Gómez, & Crook, 2013; Yang & Carless, 2013).

Son muchos los testimonios que sitúan el feedback como la más importante de las contribuciones al proceso de aprendizaje desde la perspectiva pedagógica (Black & William, 1998b; Boud & Molloy, 2013; Clark, 2012; Duncan, 2007; Evans, 2013; Gedye, 2010; Hounsell, 2003; R. McDonald et al., 2000). De hecho, destaca como la única condición referida de forma unánime en los decálogos y manifiestos sobre la calidad evaluativa consultados (tabla 35). Junto a la proacción, ocupa un lugar preferente entre los principios teóricos que orientan la evaluación para el aprendizaje, por guiarlo y orientarlo constructivamente (ARG, 2002; Carless, 2007, 2013a; Nicol et al., 2014; Stiggins, 2005a, 2005b; J. Chappuis et al., 2012; Stobart, 2010; Walker, 2013).

Dentro de la definición amplia ofrecida, el feedback puede tener lugar en modalidades diversas y complementarias. La tabla 43 presenta una propuesta de clasificación a partir de los criterios que en mayor medida aparecen reseñados en las publicaciones analizadas. En todas las tipologías resultantes cabría considerar las modalidades híbridas o mixtas, por lo que las distintas alternativas se representan con una línea divisoria de trazado discontinua.

Tabla 43
Clasificación de las principales modalidades de feedback

CRITERIOS	MODALIDADES DE FEEDBACK RESULTANTES
INTENCIONALIDAD	Formal. Aquel que el profesorado provoca deliberadamente, por medio de oportunidades estructuradas.
	Informal. Aquel que surge de forma espontánea entre las/os participantes como parte del flujo natural del proceso de enseñanza-aprendizaje.
AGENTE	Autofeedback. Es interno, procede de una/o misma/o.
	Feedback entre iguales. El intercambiado entre compañeras/os.
	Feedback docente. Aquel que ofrece el profesorado al alumnado.
	Feedback externo. Procede de agentes ajenos al contexto formativo específico de la evaluación (expertas/os, tutoras/es de prácticas, profesoras/es observadoras/es, etc.).
NATURALEZA	Respuestas modelo. Basado en respuestas ideales o ejemplares que obtendrían el 100% de la calificación. Son generadas a priori por el profesorado y distribuidas en igualdad de condiciones a todo el alumnado implicado en la evaluación. Su uso exige comparación y contraste.
	Comentarios personalizados. Comentarios de profundidad y extensión variable que responden al trabajo realizado por un equipo o estudiante en particular.
SOPORTE	Verbal. Administrado de forma oral, en conversación.
	Escrito. Administrado en soporte escrito.
CAUCE	En persona. Proporcionado cara a cara, en contacto personal.
	Documentado. Comentarios plasmados en el propio trabajo (si procede) y/o a modo de informe de extensión y profundidad variable.
	Otras- TIC. La proliferación de las tecnologías de la comunicación ofrece nuevas posibilidades didácticas e interactivas (foros, blogs, uso de plataformas, chat y video-chat, etc.). A pesar de la distancia física, éstas pueden aproximarse más a la modalidad personal o a la documentada, según el caso.
DESTINATARIO	Individual. Es personalizado, centrado en el trabajo de un individuo.
	Equipo-grupo reducido. Responde a un trabajo desarrollado en un grupo pequeño dentro del grupo-aula de referencia.
	Gran grupo. Aquel que refiere a la evolución general y el trabajo del grupo-aula de referencia, no a individuos o equipos en particular.
FOCO	Tarea. Comentarios sobre la propia tarea de aprendizaje, en términos de aspectos correctos, incorrectos o incompletos.
	Procesamiento de la tarea. Comentarios acerca de lo que la/el estudiante puede hacer para seguir adelante con una tarea de aprendizaje.
	Autoregulación. Comentarios sobre las posibilidades del alumnado para controlar y autodirigir su propio aprendizaje.
	Persona. Comentarios acerca de los atributos y rasgos personales de quien realiza la tarea valorada.

CRITERIOS	MODALIDADES DE FEEDBACK RESULTANTES
NIVEL	Específico. Es el referido a una tarea o actividad concreta.
	General. Aporta orientaciones que exceden la materia, el área de conocimiento sobre la que se está trabajando.

Nota: Elaboración propia a partir de diversas contribuciones (Clark, 2012; Evans, 2013; Ferguson, 2011; Hattie & Temperly, 2007; Huxham, 2007; Orsmond & Merry, 2010; Poulos & Mahoney, 2008; Price et al., 2010; Sambell, 2011; Sambell et al., 2007).

La calidad y efectividad de la retroalimentación no está asociada a su cantidad ni al empleo de una o varias modalidades en particular. Para contribuir a sus propósitos, su generación y administración debe comenzar librando dos dificultades clásicas, ampliamente reportadas en la literatura específica:

- *La divergencia natural de interpretaciones que se produce ante el feedback entre estudiantes y profesorado.* Docentes y discentes no comparten sus representaciones sobre lo que es el feedback (Boud & Molloy, 2013; Carless, 2006; Orsmond & Merry, 2010; Sambell, 2011). El alumnado concibe la retroalimentación como un producto que proporciona la respuesta correcta (Higgins et al., 2002; Price et al., 2010); la asocia a mensajes directos y no reconoce las aportaciones que se desmarcan del clásico patrón lineal y transmisivo, apoyado en correcciones y comentarios escritos. El profesorado, en cambio, incluye las experiencias informales en su caracterización del feedback y apela a las diversas modalidades que pueden tener lugar en soporte oral o verbal. En consecuencia, se produce el doble lamento del que dan cuenta Sambell et al. (2007): el de docentes, que no ven rentabilidad en el tiempo y el esfuerzo invertido en la retroinformación, y el de las/os estudiantes, que se manifiestan descontentos por no recibir orientaciones claras ni suficientes.
- *Las dificultades de decodificación del alumnado ante el feedback* representan otra limitación señalada de manera recurrente (Carless, 2002, 2006; J. Chappuis, 2005, 2012; Sadler, 2010). En ocasiones el alumnado no actúa ante el feedback simplemente porque no comprende los criterios a los que debe ajustarse su desempeño (Jonsson, 2013; NUS, 2010). En consecuencia, la retroalimentación debe partir de la interpretación que realicen las/os estudiantes de la tarea, el aprendizaje, el conocimiento o el lenguaje propio de la disciplina.

Este punto pretende indicar que la buena retroinformación sólo puede ser eficaz si los estudiantes y los profesores comparten las mismas ideas básicas sobre las reglas del juego. Algunas retroinformaciones buenas no

llegan a buen puerto porque el profesor no ha descubierto que los estudiantes participan en el juego universitario con reglas diferentes. (Knight, 2005, p. 196).

Frente a las dificultades reportadas, el enfoque de la evaluación para el aprendizaje exige reconceptualizar el feedback, de modo que pueda contribuir a sus finalidades pedagógicas con eficacia y solvencia (Molloy & Boud, 2013; OGG-ASKe, 2009). Fruto de la revisión de referentes realizada, el modelo ITC integra los principales rasgos hacia los que debe tender la retroinformación en ese reto, y los convierte en indicadores para este segundo criterio:

- *Basado en criterios y objetivos.* El feedback debe estar basado en la tarea, no en la persona (Black & Wiliam, 1998a, 2009) y en el contenido, no en el ego (Gibbs & Simpson, 2009). Desde esa perspectiva, la retroinformación de calidad ayuda al alumnado a interpretar las metas formativas (Atkin et al., 2001; Jessop, 2013; Laurillard, 2002; Nicol & Macfarlane, 2006; Nicol, 2007, 2009; Ramaprasad, 1983; Sadler, 1989, 2010) y subraya los aspectos centrales del aprendizaje (Duncan, 2007). Para ello, debe estar vinculado a tareas u objetivos precisos, en los que se facilite la comprensión de los criterios de evaluación (Ferguson, 2011; Price & O'Donovan, 2006; Taras, 2002).
- *Frecuente y puntual.* Para ser útil y contribuir al aprendizaje, el feedback debe ser frecuente y puntual (Cano et al., 2010; Gibbs, 2006; Huxham, 2007; Jonsson, 2013; NUS, 2010; Sambell, 2011; Sambell et al., 2007). Debe estar integrado en el proceso formativo, sin ser el fruto de una acción aislada (Boud & Molloy, 2013). Aunque la clave no está en la cantidad de información generada (Cano, 2014; S. Chappuis et al., 2009; Hattie & Timperley, 2007), para resultar efectiva la retroinformación debe llegar cuando existe margen real para la revisión, el ajuste y la mejora.
- *Descriptivo, comprensible, suficiente.* La retroinformación recibida debe ser descriptiva, ofreciendo los detalles y la profundización necesaria para facilitar su comprensión (J. Chappuis, 2005; Ferguson, 2011; Gibbs & Simpson, 2009; Orsmond & Merry, 2010). Es preciso valorar a este respecto el impacto de las dificultades de decodificación y la divergencia de concepciones e interpretaciones descritas.
- *Dialógico.* Frente al típico modelo lineal e informativo, el feedback dialógico se define a partir de “intercambios interactivos en los que las interpretaciones se comparten, los significados se negocian y las expectativas se clarifican” (Carless, 2013a, p.

117). Su desarrollo representa una experiencia de aprendizaje sensible a las necesidades informativas del alumnado (Arter, 2009; Stiggins, 2005a), a quien atribuye un papel activo (Monereo & Castelló, 2009b). Más que informar, el feedback dialógico procura guiar a las/os estudiantes (Sadler, 2010), ofreciéndoles la oportunidad de plantear abiertamente sus dudas y desacuerdos ante el mensaje recibido (Gibbs, 2006; Laurillard, 2002). Les facilita la oportunidad de acercarse e interactuar con los estándares de calidad de la disciplina, sin cuya decodificación no sería posible comprender el nivel de logro evidenciado en su trabajo (Harlen, 2003; Harrington et al., 2006; Hounsell, 2003, 2007). Por todo ello, se convierte en un recurso para la autorregulación del aprendizaje discente, estimulando la reflexión sobre las primeras producciones y el empleo de las conclusiones en actividades futuras (Cano, 2014). En consecuencia, ante las dificultades de interpretación reseñadas, el feedback dialógico ofrece mayores garantías de resultar útil y comprensible a las/os destinatarias/os (Carless, 2013b; Carless et al., 2011; OGG-ASKe, 2009; Orsmond et al., 2013; Sambell, 2011).

De la misma forma que el aprendizaje no se produce por la mera transmisión de información escrita o hablada, la provisión de feedback no guía en sí misma la mejora del aprendizaje. Para que los estudiantes aprendan deben hacer algo con la información transmitida, analizar el mensaje, hacer preguntas sobre la misma, discutirla con otras personas, conectarla con los conocimientos previos y emplearla para cambiar acciones futuras. Lo mismo es aplicable para los comentarios de feedback. Si bien la calidad de los comentarios, es importante, la calidad de la interacción de los alumnos con estos comentarios es igual de importante, o tal vez más. (Nicol, 2010, p. 503).

- *Sostenible.* La reconceptualización del feedback en la evaluación para el aprendizaje supone la incorporación de nuevos agentes y modalidades de intervención al mismo (Nicol et al., 2014; Sambell, 2011; Sambell et al., 2007; Yorke, 2003). Admitida la relevancia y necesidad de retroalimentación, su administración exclusiva a cargo del profesorado no resultaría sostenible (Ballantyne et al., 2002; Brown, 2003a; Carless et al., 2011; Hounsell, 2007; Liu & Carless, 2006; Lu & Law, 2012). Así por ejemplo,

el prestigioso modelo de evaluación del Alverno College de Milwaukee⁵⁵ (EEUU), una de forma indisoluble evaluación y feedback, pudiendo proceder este último del profesorado, de compañeras/os y de otras/os profesionales. De forma más precisa, Carless (2013a) lo define como la “participación activa de los estudiantes en actividades en las que generan y usan feedback propio, de compañeros o de otros en un proceso continuo de desarrollo de sus capacidades como aprendices autónomos” (p. 118). Los beneficios pedagógicos de dicha participación se encuentran ampliamente suscritos (I. Álvarez & Olivera-Smith, 2014; I. Álvarez, 2008a; Boud et al., 1999; Dochy et al., 1999; Falchikov & Goldfinch, 2000; Liu & Carless, 2006; Sanmartí, 2007, 2010) y se asocian a la generación de procesos de autofeedback o de feedback entre iguales.

- *Estimulante.* En la evaluación del aprendizaje coexisten dos dimensiones relevantes con incidencia directa sobre el alumnado y sobre el personal docente: la dimensión técnica y la emocional (Boud, 1995; Falchikov & Boud, 2007). Yang y Carless (2013) consideran que el componente emocional y motivacional representa uno de los vértices del triángulo de la generación y provisión de feedback, junto al propio contenido y al proceso comunicativo. Su efecto puede ser positivo o negativo, dependiendo de su contenido, de la forma de administrarlo y de las expectativas del alumnado al respecto (Higgins et al., 2002). En consecuencia, es importante que la retroinformación no se limite a medir el fracaso (Wootton, 2002) y promueva mensajes motivacionales positivos (J. Chappuis et al., 2012; Nicol, 2007, 2009; Stiggins, 2007). También que tome distancia de imposiciones autoritarias, ya que no resultan constructivas para el aprendizaje (Jonsson, 2013).
- *Prospectivo.* El feedback debe ayudar al alumnado a comprender no sólo dónde se encuentra, sino lo que debe hacer para avanzar hacia las metas definidas (Carless, 2013a; Ramaprasad, 1983; Stiggins, 2008; Walker, 2013). Debe centrarse en la mejora y ofrecer oportunidades de cambio y avance (Gibbs & Simpson, 2009; B. Salinas & Cotillas, 2007; Wootton 2002). Sobre ello versa la proacción, segundo nivel de profundización en los resultados evidenciados, que se aborda en los dos últimos criterios.

⁵⁵ El sitio web del centro (<http://www.alverno.edu/grades/>) información completa al respecto.

2.5.3 Proacción para la mejora del aprendizaje.

El carácter prospectivo del feedback da lugar a la *proacción*, entendida como la capacidad para determinar pautas claras de actuación, ajuste y mejora que contribuyan al aprendizaje a partir de los resultados constatados. Del inglés *feedforward*, Castillo (2004), Toledo (2006) o Sabirón & Arraiz (2013) denominan *retroacción* a este segundo nivel de profundización. Igual que en G. Rodríguez (1998), en la investigación en curso el término ha sido traducido como *proacción*, valorando la mayor pertinencia de los matices del prefijo *pro*⁵⁶ para el criterio calificado.

Desde el enfoque teórico de la evaluación para el aprendizaje, se reivindica la necesidad de enfocar el feedback al aprendizaje y no a la mera calificación (Gibbs, 2006). Para ello, debe adoptar carácter prospectivo e integrar propuestas de actuación que contribuyan a reducir la distancia entre los resultados esperados y los constatados (Hattie & Timperley, 2007; Hounsell, 2003; Nicol, 2013; Orsmond et al., 2013; Taras, 2002).

Además de ayudar a comprender las metas de aprendizaje a las que debe aproximarse y su situación actual respecto de las mismas, la retroinformación debe incorporar posibles actuaciones para comenzar el avance (Atkin et al., 2001; Jonsson, 2013; Ramaprasad, 1983; Walker, 2013). Entre otras muchas alternativas, éstas pueden sugerir retomar la misma tarea (repetición, revisión, corrección), profundizar en ella o desarrollar nuevas actividades relacionadas.

Desde esta perspectiva, Stiggins (2008) concluye que el feedback sólo resulta motivador y efectivo para el alumnado bajo la confluencia de las siguientes condiciones: cuando resulta comprensible, cuando proporciona pautas de actuación para próximas intervenciones y cuando plantee una mejora que parezca realizable y anime a seguir trabajando. Igual que en la retroalimentación, Jonsson (2013) advierte que el alumnado presenta dificultades para hacer uso efectivo de las recomendaciones y orientaciones derivadas de la evaluación; efectuado un análisis de más de 100 estudios sobre el tema, concluye que éstas pueden relacionarse con la incomprensión de la jerga académica o con la falta de estrategias para su uso productivo.

⁵⁶ Según el *Diccionario de la Lengua Española* (Real Academia Española, 2014), denota impulso hacia delante y en favor de.

2.5.4 Proacción o regulación de la enseñanza.

Como afirma Villar (2009),

la profesión de enseñante tiene en la evaluación una de las tareas más prominentes y simultáneamente más determinante para el progreso de los estudiantes y la reflexión docente, porque tres son los objetos de una evaluación: uno es el aparentemente estático y terminal del aprendizaje del alumnado; otro es más dinámico, debiera ser auténtico y muchas veces se escamotea: nos referimos a los procesos de enseñanza; y finalmente, el vaporoso e incommensurable de la práctica docente. (p. 3).

Esta visión remite al concepto de *scholarship of teaching and learning* y presenta la figura docente como “[...] profesional que asume la reflexión sobre la acción, como pilar de su ejercicio profesional, investigando sobre su propia actividad docente” (Chocarro et al., 2013). Concebida como instrumento de indagación, la evaluación ayuda al profesorado a comprender el proceso de aprendizaje por el que transcurre el alumnado y las concepciones erróneas o certeras que al respecto sostiene (Ramsden, 2003). A resultados de la misma, pueden plantearse medidas de proacción sobre la docencia para mejorarla y reforzar el aprendizaje discente (Castillo, 2004; Valcárcel, 2003).

La buena evaluación refuerza la instrucción e incide en la enseñanza (Birembaum, 1996; Camilloni, 1998; McMillan, 2000), representando un recurso esencial para el desarrollo institucional y docente (Boud & asociados, 2010; CRESST, 2012; Gibbs, 2010; Stiggins & Duke, 2008). Así lo entienden también O. Esteve (2010), Sanmartí (2007) o Bain (2007), señalando la necesidad de que los docentes realicen una doble lectura de las valoraciones y calificaciones discentes, que les permita revisar y mejorar su enseñanza y no sólo la actuación del alumnado.

Si queremos promover la construcción de una comunidad de aprendizaje -donde los estudiantes buscan naturalmente la retroalimentación crítica de su propio trabajo- sería razonable que el profesorado modele este mismo compromiso, empleando sistemáticamente los datos en referencia a su proceso de enseñanza y aprendizaje. (Shepard, 2000, p. 12).

Desde esta perspectiva, el feedback debe favorecer la reflexión sobre la docencia y contribuir a adecuarla en favor del aprendizaje (Nicol, 2007, 2009; Nicol & Macfarlane, 2006).

2.6 Revisión del proceso de evaluación: metaevaluación.

La revisión de la evaluación o metaevaluación parte de reconocer la imposibilidad de llevar a cabo prácticas evaluativas perfectas (Jorba & Sanmartí, 2000; McMillan, 2000).

El objetivo de la metaevaluación es asegurar la calidad de los servicios evaluativos, evitar o enfrentarse a las prácticas ilegales o a los servicios que no son de interés público, señalar el camino para el perfeccionamiento de la profesión y promover una mayor comprensión de la empresa evaluativa. (Stufflebeam & Shinkfield, 1987, p. 52).

En el modelo ITC su desarrollo se plantea de acuerdo a dos criterios interrelacionados:

2.6.1 Revisión específica del dispositivo evaluativo.

La necesidad de llevar a cabo una revisión sistemática y crítica de la evaluación, que permita identificar sus puntos fuertes y sus áreas de mejora, se encuentra ampliamente representada entre los distintos tipos de referentes consultados. Es uno de los principios de calidad explícitamente recogido en las orientaciones teóricas de Cano et al., (2010), Brown et al. (1996) y Boud y asociados (2010) y subyace a las propuestas estructuradas o estandarizadas que formulan directrices para el diseño y revisión del proceso evaluativo (Astin et al., 1992; Baker et al., 1996, ENQA, 2009; JCSEE, 2007; Land, 1997; QAA, 2014).

La metaevaluación es reconocida como un medio eficaz para valorar la calidad de la evaluación y contribuir a ella (Pattel, 2002; Stake, 2006; Stufflebeam, 2001b). Sin embargo, situados en el enfoque teórico de la evaluación para el aprendizaje, su objetivo también debe ser la promoción del aprendizaje discente, aunque por medio de un itinerario más amplio. La revisión ha de proporcionar elementos para mejorar la evaluación, favoreciendo

con ello una estimulación más eficaz del aprendizaje y una acreditación más precisa y adecuada del mismo.

En consecuencia, el modelo ITC propone una revisión que va más allá del análisis técnico de la evaluación y se concreta en tres dimensiones:

- *Dimensión pedagógica.* Esta primera indagación conduce a la revisión del principio de consecuencialidad, que determina hasta qué punto representa el sistema evaluativo un recurso eficaz para impulsar el trabajo y el aprendizaje de las/os estudiantes (Gielen et al., 2003; Hargreaves, 2007; Monereo & Castelló, 2009a, 2009b). Supone el análisis de su potencialidad didáctica (Cano & Ion, 2012). Rotger (1992) señala en este mismo sentido que la evaluación debe responder al principio de utilidad⁵⁷, de forma que las orientaciones que de ella se desprendan contribuyan a mejorar el proceso y el resultado de aprendizaje.
- *Dimensión técnica.* La evaluación representa un proceso complejo desde el punto de vista técnico (Margalef, 2009; Rué, 2007, 2008; D. Salinas, 2002). En consecuencia, la metaevaluación debe reparar en la satisfacción de aquellos requisitos metodológicos que comprometan la calidad de las conclusiones alcanzadas. Entre los que en mayor medida aparecen reseñados en la literatura, destacan los siguientes:
 - ✓ Validez. El principio de validez determina de forma general si el sistema evaluativo valora realmente el aprendizaje que pretende, sin desviaciones ni simplificaciones que lo distancien del mismo (Baker et al., 1996; Monereo & Castelló, 2009a; Ortiz, 2006; Pérez Pérez, 2004).
 - ✓ Fiabilidad. El principio de fiabilidad determina la confiabilidad del sistema evaluativo para valorar con precisión los aprendizajes previstos (Monereo & Castelló, 2009b; Ortiz, 2006; Pérez Pérez, 2004).
 - ✓ Viabilidad. El principio de viabilidad determina que el sistema evaluativo resulte aplicable en las condiciones reales de su puesta en práctica (Care, 1978; Rotger, 1992). Dicho principio se corresponde con los estándares de viabilidad del JCSEE (2007) y conecta con la inquietud por la sostenibilidad de las prácticas evaluativas (Buscà, Capllonch et al., 2010; Cano et al., 2010; Suskie, 2006).

⁵⁷ El empleo del término utilidad aquí no se corresponde estrictamente con la definición planteada para los estándares de utilidad del JCSEE (2007) o del CRESST (Baker et al., 1996; Land, 1997), que es más amplia y no se limita a la finalidad pedagógica (satisfacción de necesidades informativas o propósitos).

- ✓ **Sistematicidad.** El principio de sistematicidad está relacionado con la premeditación de la evaluación (Pérez Pérez, 2004); hace referencia a la intencionalidad y la planificación, de modo que el sistema evaluativo se diseñe y aplique de forma controlada y ordenada.
- **Dimensión ética.** La ética representa una exigencia fundamental a la que todo proceso de evaluación debería responder con diligencia y responsabilidad (Batalloso, 2004; Buscà, Capllonch et al., 2010; Cano et al. 2010; A. Calatayud, 2008; McDowell & Sambell, 2003; McMillan, 2000; Race, 2003; Zabalza, 2003). La evaluación debe ser respetuosa con la persona, lo que exige reflexionar sobre una serie de principios básicos, ampliamente reseñados:
 - ✓ **Justicia.** Aunque aparece asociado a definiciones y acepciones diversas, se trata de un principio aceptado de forma general en la literatura. Sobre el mismo pueden discriminarse dos grandes interpretaciones: una primera visión que plantea la justicia en términos compensatorios (M. Medina & Verdejo, 2000; Baker et al., 1996) y una segunda que la vincula a la ecuanimidad, a la igualdad de condiciones y a la no arbitrariedad (Cano et al., 2010; Monereo & Castelló, 2009a). El modelo ITC se inclina hacia esta última, adoptando como criterio de elección la dirección que emanan los referentes normativos, más orientados a la universalidad de la evaluación que a su diversificación (ANECA, AQU, & ACSUC, 2007b; ENQA, 2009; Real Decreto 1791/2010). En consecuencia, insiste fundamentalmente en la no arbitrariedad en su ejercicio (Cano et al., 2010).
 - ✓ **Privacidad.** El principio de privacidad refiere a la salvaguarda de los derechos de anonimato del alumnado en la evaluación, así como a la confidencialidad de la información académica generada. Insta a garantizar la protección de los datos, resultados e informaciones recopiladas de las/os estudiantes (Buscà, Capllonch et al., 2010; M. Medina & Verdejo, 2000). Se trata de otro principio admitido de forma general, que encuentra un especial reflejo en las propuestas estandarizadas (Baker et al., 1996; JCSEE, 2007; Land, 1997).
 - ✓ **Beneficencia⁵⁸.** El principio de beneficencia determina que la evaluación se desarrolle libre de maleficencia, es decir, en beneficio de las/os estudiantes y sin ocasionarles perjuicio o daño.

⁵⁸ Recoge la fusión de varios principios enumerados en los referentes consultados: el principio ético (Buscà, Capllonch et al., 2010), los de beneficencia y no maleficencia (M. Medina & Verdejo, 2000) o la amistosidad (Cano et al., 2010).

- ✓ **Transparencia.** Este cuarto principio hace referencia a los derechos de información explícita, clara y transparente sobre la evaluación que se reconocen al alumnado en los referentes normativos (ANECA, AQU, & ACSUC, 2007b; ENQA, 2009; Real Decreto 1791/2010).

La descripción planteada para cada una de las dimensiones no representa sino una guía abierta para orientar la fase de revisión o metaevaluación. Los principios éticos, técnicos y pedagógicos sugeridos representan referencias admitidas de común acuerdo en las síntesis revisadas y son acordes a la definición de calidad evaluativa que subyace al modelo ITC. No obstante, la determinación final de las dimensiones y principios a considerar en cada caso, debería determinarse atendiendo a las pretensiones y prioridades que motiven la metaevaluación.

En cuanto a los agentes responsables de la revisión, caben distintas alternativas. El nivel de participación más elemental en la metaevaluación es el que tiene lugar a cargo del propio profesorado, cuando revisa la evaluación realizada para mejorarla. Stake (2006) lo define como una “supervisión autocorrectora que controla los errores que se van produciendo a lo largo de la evaluación” (p. 69). Sin embargo, la consideración de diferentes perspectivas representa un factor deseable en cualquier proceso evaluativo, incluida la metaevaluación. En consecuencia, el ITC establece la conveniencia de contar con la participación de diversos agentes en el proceso de revisión. Así lo sostiene también B. McDonald (2010) apoyada en el modelo 360º de evaluación.

En la formación universitaria resulta interesante la implicación del alumnado en el proceso de metaevaluación, especialmente en los grados vinculados a la formación básica de maestras/os, donde las/os estudiantes cuentan con herramientas para llevar a cabo una valoración semi-cualificada (Bilbao, 2011b). Además de enriquecer la revisión aportando otra perspectiva, su intervención puede dar lugar a procesos de reflexión que contribuyan al desarrollo profesional del profesorado (Capllonch et al., 2009; V. M. López et al., 2007; V. M. López, 2009a, 2011). Para las/os futuras/os evaluadoras/es, representa asimismo una oportunidad para formarse un juicio profesional sobre la complejidad de una función que deberán ejercer (Ureña et al., 2009).

Docentes y discentes representan las dos audiencias más directamente involucradas en la evaluación, pero el proceso de revisión podría reforzarse con la intervención de otros agentes según los propósitos y las posibilidades particulares. Entre ellos se encuentran las/os colegas o superiores de la propia institución y las/os asesoras/es externas/os (Capllonch et al., 2009; Santos, 2003; Stake, 2006; Tejedor, 2003). También podría recurrirse a estudiantes egresadas/os (Berk, 2005; Hamodi & López, 2012; P. González &

Grande, 1999) o agentes empleadores (Boud & asociados, 2010). En cualquiera de los casos, su colaboración podría servir a la recopilación de evidencias, a su análisis y/o a la discusión y valoración de las posibles alternativas de mejora.

2.6.2 Proacción para la mejora de la evaluación.

En coherencia con la conceptualización general propuesta en la presente investigación para el proceso evaluativo, la realización de una metaevaluación no se justifica por sí misma. Su objetivo debe ser facilitar la comprensión del sistema o sistemas de evaluación que evalúa, para poder emprender actuaciones pedagógicas o acreditativas en consecuencia. Como resume Sime (1998), la metaevaluación debe mirar a la evaluación para volver a ella.

En el contexto específico del modelo ITC, la revisión de las dimensiones ética, técnica y pedagógica de la evaluación debe dar lugar a dos productos complementarios. Por una parte, debe permitir identificar con rigor y realismo los límites y posibilidades del dispositivo evaluativo articulado, en especial en aquellos aspectos que afecten a la calificación del aprendizaje discente. Por la otra, tal valoración debería dar lugar a decisiones y actuaciones para contribuir a una evaluación de mayor calidad (revisión y redefinición de tareas y condiciones, introducción de ajustes, neutralización de riesgos, etc.). Aunque la metaevaluación también pueda adoptar propósitos acreditativos (B. McDonald, 2010; Monereo & Castelló, 2009b; Stufflebeam, 2001b) esta segunda alternativa no se contempla en la investigación en curso.

2.7 Modelado en la formación básica de maestras/os de Educación Infantil o Primaria.

La teoría de la evaluación declarada en el modelo ITC adopta la formación básica de maestras/os de Educación Infantil o Primaria como referencia inmediata y a ella dirige la propuesta de criterios e indicadores en cada subproceso. Aun así, las orientaciones ofrecidas para la praxis evaluativa son aplicables a otras titulaciones universitarias, si se efectúa un trabajo previo de adecuación conceptual y terminológica. No sucede lo mismo en este último subproceso, que plantea requisitos de aplicación exclusiva en los grados vinculados a la preparación de futuras/os evaluadores.

De acuerdo con Fullan (1991) en la enseñanza es importante predicar con el ejemplo; a ello responde este último subproceso, vinculado al modelado consciente de prácticas evaluativas de calidad. Su presencia y representación entre los referentes consultados es mucho menor que la constatada en los subprocesos anteriores e indirecta. Su incorporación al modelo ITC surge al advertir que en la formación básica de maestras/os el alumnado debe desarrollar competencias ligadas a la evaluación. El desarrollo de las mismas no sólo depende de la teoría que se explicita en los diversos módulos y asignaturas, ya que la vivencia evaluativa directa tiene un impacto mucho mayor en la generación de representaciones (Campbell, 1993/2012; J. Contreras, 2010; V. M. López, 2004; Marín & Cervi, 2006). De acuerdo con este planteamiento, son varios los testimonios que advierten en las/os maestras/os en ejercicio la reproducción de los patrones evaluativos experimentados (B. Álvarez et al., 2007; Hamodi & López, 2012; Tejedor, 1999), incluso cuando éstos no concuerden con la teoría explícita sostenida.

En consecuencia, en la formación básica de maestras/os la evaluación del aprendizaje debe representar un ejemplo práctico de los principios y orientaciones teóricas ofrecidas. Ésta debe ser coherente, de modo que no se incurra en contradicciones entre lo que se dice en evaluación y lo que verdaderamente se hace (Palacios & López, 2013).

A juicio de Stiggins (1998) la evaluación debe representar una experiencia instruccional, es decir, plenamente integrada en el proceso formativo y favorecedora de la consecución de las metas de aprendizaje previstas, incluidas las referentes a la capacitación evaluativa. Ello exige disponer deliberadamente de situaciones de evaluación participativa para estimular el aprendizaje y trabajo del alumnado en este sentido. Además de propiciar una ocasión para poner en práctica las directrices teóricas (evaluación práctica y contextualizada), estas situaciones revelan de manera evidente los retos y dificultades asociadas a las demandas pedagógicas y/o acreditativas de evaluación (I. Álvarez, 2008; C. Brew et al., 2009; V. M. López et al., 2012; McGarr & Clifford, 2013).

3. Modelo ITC

El nuevo apartado expone completo el modelo de evaluación del aprendizaje asumido como referencia en la presente investigación. Recurriendo a un formato de ficha que facilita su lectura, cada subproceso se presenta acompañado de los criterios e indicadores anticipados para su desarrollo en el **ANEXO 04:**

3.1 Subproceso 01: planificación consciente e integrada del proceso de evaluación.

Definido a continuación, el primero de los subprocesos identificados incorpora en su definición seis criterios y un total de 32 indicadores para su desarrollo (ocho son repetidos).

1. PLANIFICACIÓN CONSCIENTE E INTEGRADA DEL PROCESO DE EVALUACIÓN.

El primer subproceso determina la necesidad de una evaluación del aprendizaje plenamente integrada en la actividad docente y discente. Presente desde el comienzo del proceso formativo, su delimitación debe partir de la consideración técnica consciente de los siguientes elementos (criterios).

1.1. El dispositivo evaluativo representa un recurso estratégico para la estimulación intencional y consciente de los aprendizajes previstos.

Este primer criterio centra la atención en la potencialidad estratégica del dispositivo evaluativo articulado, entendida como su capacidad para promover deliberadamente los aprendizajes previstos, y no sólo medirlos y acreditarlos. Basado en los efectos que suscita la evaluación sobre el esfuerzo y el trabajo del alumnado, integra en la planificación evaluativa los principios y condiciones favorecedoras del aprendizaje, señalados desde la literatura específica.

El criterio se concreta en los siguientes ocho rasgos evaluativos o indicadores, entre otros posibles:

1.1.1. Evaluación integrada.

Cuidadosamente delimitado y planificado desde el comienzo, el sistema de evaluación está plenamente integrado en el proceso de enseñanza y aprendizaje, de modo que ni es independiente, ni llega al final.

1.1.2. Evaluación continua (consultar también 1.1.4.).

El sistema de evaluación articulado prevé la realización de diferentes actividades evaluativas en los distintos momentos del proceso de formativo. Ello se traduce en el desarrollo de una dinámica evaluativa inicial, procesual y final, que combine de forma equilibrada las funcionalidades acreditativas y pedagógicas de la evaluación e integre la valoración de los procesos y los productos significativos y relevantes a tal efecto.

1.1.3. Gestión eficaz y adecuada del tiempo de trabajo y esfuerzo (consultar también 1.1.4.).

El dispositivo evaluativo está diseñado para impulsar de forma ajustada y deliberada en el alumnado la cantidad y el tipo de trabajo y esfuerzo requerido para la consecución de las metas de aprendizaje.

1.1.4. Evaluación sensible con el proceso de aprendizaje.

La evaluación es sensible al proceso de construcción vivenciado por la/el estudiante en el aprendizaje y no lo perjudica con demandas o exigencias excesivas. Está presente en las distintas fases del proceso formativo e impulsa el trabajo y el esfuerzo necesarios, sin someter al alumnado a una presión o a un ritmo de trabajo poco ajustados a su evolución y sus necesidades.

1.1.5. Evaluación explícita (ver 2.1.1.).

1.1.6. Evaluación alineada (ver 1.2.2.).

1.1.7. Evaluación formativa o reguladora (consultar también 1.1.8.).

El dispositivo evaluativo incorpora actividades que potencian el uso pedagógico y formativo de la evaluación. Realizadas con finalidades preventivas, diagnósticas, de comprensión y de mejora, estas actividades conducen al análisis de los logros y fortalezas y de las deficiencias y errores en el proceso de aprendizaje, para efectuar la regulación consiguiente del mismo bajo dirección docente.

1.1.8. Evaluación formadora o autorreguladora.

El dispositivo evaluativo incorpora actividades que potencian el uso pedagógico y formador de la evaluación. Realizadas con finalidades preventivas, diagnósticas, de comprensión y de mejora, estas actividades conducen al propio alumnado al análisis de sus logros y fortalezas y de las deficiencias y errores en el proceso de aprendizaje para proceder a su autorregulación autónoma.

1.2. El dispositivo evaluativo se diseña garantizando los principios de alineación y coherencia.

Este segundo criterio centra la atención en la relación de alineación y coherencia que debe preservar la planificación evaluativa con respecto al plan de estudios y los componentes curriculares que lo delimitan. Se trata de garantizar la integración de la evaluación en la programación de un módulo o asignatura desarrollada a corto plazo, sin desligarse de las expectativas o metas específicas profesionales determinadas a largo plazo en la titulación de referencia.

El criterio se concreta en los dos siguientes rasgos evaluativos o indicadores, entre otros posibles:

1.2.1. Alineación de la evaluación con el plan de estudios.

El sistema de evaluación definido valora aprendizajes coherentes con los logros profesionales definidos a largo plazo en el plan de estudios de referencia.

1.2.2. Alineación de la evaluación con el resto de componentes curriculares.

El sistema de evaluación definido muestra coherencia con el resto de componentes que conforman el currículo, en especial, con los objetivos de aprendizaje previstos y la estrategia metodológica delimitada para su consecución (actividades de aprendizaje desarrolladas).

1.3. El proceso de evaluación se realiza con herramientas y procedimientos seleccionados de forma consciente y justificada.

Este criterio centra la atención en la elección metodológica que debe tener lugar en la fase de planificación de la evaluación. Dado que no existen procedimientos evaluativos neutrales ni perfectos, se describe en él la compleja elección que debe tener lugar en cada caso, valorando de forma justificada la adecuación de las tareas y herramientas dispuestas, el repertorio de evidencias de aprendizaje que pueden ofrecer y su impacto en la validez y fiabilidad del proceso evaluativo.

El criterio se concreta en los seis siguientes rasgos evaluativos o indicadores, entre otros posibles:

1.3.1. Adecuación pedagógica.

Los procedimientos, tareas y herramientas empleadas en la evaluación son seleccionados de forma consciente y justificada de acuerdo a razones de naturaleza pedagógica, vinculadas al tipo de aprendizaje que se pretende valorar y promover y a las evidencias que para ello se precise recopilar.

1.3.2. Adecuación al contexto formativo.

Los procedimientos, herramientas y tareas previstas para la recogida de evidencias de aprendizaje se eligen por su adecuación a las características y exigencias que plantea el contexto formativo de referencia (la educación superior y sus condiciones prácticas en cuanto a tamaño del grupo, espacios, tiempos, etc.).

1.3.3. Adecuación al perfil discente.

Los procedimientos, herramientas y tareas previstas para la recogida de evidencias de aprendizaje se eligen por su adecuación a las características y las particularidades del perfil discente (personas adultas en proceso de formación profesional básica, con necesidades y demandas particulares).

1.3.4. Variedad de herramientas y procedimientos (consultar también 1.3.6. y 1.5.7.).

El dispositivo evaluativo prevé la combinación de técnicas y procedimientos diferentes de evaluación, de forma que se contrarresten las limitaciones parejas a cada uno de ellos, se incremente el rigor de la valoración y se abarquen y reporten evidencias sobre los distintos aspectos implicados en el aprendizaje.

1.3.5. Validez de las herramientas y procedimientos evaluativos (consultar también 4.1.2. y 6.1.2.).

El dispositivo evaluativo incorpora tareas, técnicas y procedimientos válidos para evaluar los aprendizajes previstos en el alumnado universitario. Las evidencias recopiladas reflejan diferentes componentes o aplicaciones de dicho aprendizaje, no aspectos circundantes o relacionados.

1.3.6. Fiabilidad de las herramientas y procedimientos evaluativos (consultar también 4.1.1. y 6.1.2.).

Las tareas, técnicas y herramientas de evaluación previstas proporcionan una valoración precisa y rigurosa del aprendizaje discente. Captan y discriminan diferentes niveles de consecución según el logro evidenciado en los criterios de evaluación y calificación establecidos (consultar también 2.1.4.).

1.4. El sistema de evaluación planificado articula medidas efectivas para responder a la diversidad natural de necesidades, capacidades y trayectorias presentes en el grupo de estudiantes.

El cuarto criterio centra la atención en la capacidad del dispositivo evaluativo para ofrecer respuesta efectiva a las necesidades, trayectorias, capacidades y talentos desiguales presentes en un mismo grupo de estudiantes. Sin contravenir las exigencias de la evaluación criterial en el marco actual de la formación universitaria, plantea la articulación de medidas que garanticen mecanismos no estereotipados para la valoración y promoción del aprendizaje.

El criterio se concreta en los dos siguientes rasgos evaluativos o indicadores, entre otros posibles:

1.4.1. Flexibilidad y elección (consultar también 1.5.7.).

El dispositivo evaluativo ofrece alternativas diferentes para que las/os estudiantes, desde su particular situación de partida y sus aspiraciones, desarrollen y demuestren los aprendizajes vinculados a las metas formativas definidas (sistemas de calificación paralelos, actividades diferentes, distintos plazos de realización, demostración o entrega, etc.).

1.4.2. Negociación y consenso (consultar también 3.2.1.).

El dispositivo evaluativo se muestra permeable a la negociación de las condiciones operativas para su desarrollo (instrumentos, plazos, o condiciones prácticas) en aras a favorecer un mejor ajuste a la realidad particular de cada estudiante (talentos, experiencias, y circunstancias y necesidades personales) y con ello el desarrollo óptimo de las competencias definidas.

1.5. El proceso de evaluación se centra en el aprendizaje de competencias.

Este criterio plantea el resultado de la evolución de un modelo formativo basado en la enseñanza y los contenidos al nuevo paradigma basado en el aprendizaje de competencias. Describe sistemas de evaluación que asumen el aprendizaje discente como primera referencia, entendiendo que éste se define y operativiza en términos de competencia.

El criterio se concreta en los nueve siguientes rasgos evaluativos o indicadores, entre otros posibles:

1.5.1. Evaluación globalizada.

El dispositivo evaluativo plantea la recogida de evidencias de aprendizaje por medio de tareas o situaciones globalizadas, en las que se exige la integración de conocimientos, procedimientos y actitudes desde una perspectiva interdisciplinar. Evaluación holística, multidimensional o comprensiva.

1.5.2. Evaluación contextualizada.

El dispositivo evaluativo plantea la recogida de evidencias de aprendizaje a partir de tareas o situaciones reales (evaluación auténtica) o verosímiles (evaluación de la ejecución).

1.5.3. Evaluación práctica.

El dispositivo evaluativo plantea la recogida de evidencias de aprendizaje a partir de tareas o situaciones prácticas que, sin limitarse a demandas teóricas, exigen actuación discente, aplicación y transferencia.

1.5.4. Evaluación continua (ver 1.1.2.).

1.5.5. Orientación y seguimiento individual (ver 3.2.1.).

1.5.6. Evaluación explícita (ver 2.1.1.).

1.5.7. Evaluación diversa.

El dispositivo evaluativo plantea la recogida de evidencias de aprendizaje en situaciones diversas, combinando la aplicación de procedimientos y herramientas evaluativas diferentes, que den acceso a pruebas y manifestaciones de distinta índole y respondan a la diversidad natural de estilos de trabajo y aprendizaje.

1.5.8. Evaluación divergente.

El dispositivo evaluativo plantea la recogida de evidencias de aprendizaje en situaciones complejas (poco estructuradas), que admiten distintas perspectivas de análisis y resolución.

1.5.9. Evaluación participativa o compartida (ver 1.6.1.).

1.6. Evaluación participativa.

Este sexto criterio hace referencia a la previsión de los agentes intervinientes en la evaluación. Esta participación puede afectar al desarrollo de las cuatro fases del proceso evaluativo, total o parcialmente, y adoptar un desigual grado de repercusión según se establezca en cada caso. Aunque no es la única que puede tener lugar, en la enseñanza universitaria en general y, muy especialmente, en la formación básica de maestras/os, la implicación del alumnado resulta especialmente relevante.

El criterio se concreta en los cinco siguientes rasgos evaluativos o indicadores, entre otros posibles:

1.6.1. Evaluación participativa o compartida.

El dispositivo evaluativo prevé la intervención de distintos agentes de evaluación con responsabilidades y funciones pedagógicas y/o acreditativas bien delimitadas para cada caso.

1.6.2. Negociación y consenso (ver 1.4.2.).

1.6.3. Feedback sostenible (ver 5.2.8.).

1.6.4. Posicionamiento docente (ver 7.1.3.).

1.6.5. Metaevaluación participativa (ver 6.1.4.).

3.2 Subproceso 02: comunicación y difusión de criterios y condiciones.

Definido a continuación, el segundo subproceso se concreta en dos criterios y un total de cinco indicadores para su desarrollo.

2. COMUNICACIÓN Y DIFUSIÓN DE CRITERIOS Y CONDICIONES.

El segundo subproceso sienta las bases para una evaluación del aprendizaje explícita y transparente, incorporando acciones encaminadas a garantizar la comunicación y difusión de los criterios y condiciones básicas de la misma entre las/os estudiantes y cualquier otra audiencia interesada. Presente en las fases de planificación e implementación de la evaluación, este subproceso conlleva la consideración técnica consciente de los siguientes elementos (criterios).

2.1. El proceso de evaluación parte de la explicitación y difusión de las principales condiciones y criterios que habrán de condicionar su desarrollo.

Este primer criterio centra su atención en la explicitación y difusión de las condiciones básicas que caracterizan el desarrollo del proceso evaluativo en la práctica (plazos, procedimientos, instrumentos, criterios, calificación, etc.). Considerada la repercusión de la evaluación sobre el trabajo y el aprendizaje de las/os estudiantes, este criterio

combate de forma deliberada los espacios de indeterminación que puedan surgir en la fase de planificación e informa de las cuestiones centrales que habrán de condicionar la valoración del aprendizaje al comienzo de la implementación y durante la misma.

El criterio se concreta en los cuatro siguientes rasgos evaluativos o indicadores, entre otros posibles:

2.1.1. Evaluación explícita.

Las distintas decisiones que dan lugar al sistema de evaluación (criterios, plazos, procedimientos, instrumentos, condiciones, etc.) se encuentran debidamente explicitadas a fin de garantizar los principios de publicidad y transparencia y su repercusión consiguiente en el proceso de aprendizaje discente.

2.1.2. Negociación y consenso (ver 1.4.2.).

2.1.3. Expectativas de aprendizaje elevadas.

Los criterios de evaluación comunican expectativas de aprendizaje elevadas pero razonables, en conexión directa con las competencias y estándares específicos de la futura profesión.

2.1.4. Establecimiento de niveles de logro/dificultad diferentes (consultar también 4.2.2.).

Los criterios de evaluación definen diferentes niveles de dificultad o logro con relación al aprendizaje a efectuar, ofreciendo al alumnado una perspectiva de la progresión que habrá de protagonizar en el proceso de desarrollo de las competencias.

2.2. En su desarrollo el dispositivo evaluativo incorpora medidas para favorecer la comprensión e interiorización de sus criterios y condiciones básicas por parte del alumnado.

Este segundo criterio pone el acento en el proceso de clarificación que debe tener lugar en la implementación del sistema de evaluación para favorecer la comprensión profunda y la interiorización de los estándares y criterios en los que se apoya la valoración del aprendizaje.

El criterio se concreta en el siguiente rasgo evaluativo o indicador, entre otros posibles:

2.2.1. Interiorización y ajuste de representaciones (consultar también 5.2.1.).

El dispositivo evaluativo incorpora actividades ideadas para profundizar en la comprensión de los estándares a los que ha de aproximarse el aprendizaje y a los criterios que rigen su valoración. Dichas actividades pueden desarrollarse con la implicación directa del alumnado como agente evaluador (consultar también 1.6.1. y 1.4.2.) o sin ella.

3.3 Subproceso 03: recopilación y registro de las evidencias de aprendizaje.

Definido a continuación, el tercer subproceso se concreta en dos criterios y cuatro indicadores que los desarrollan.

3. RECOPIACIÓN Y REGISTRO DE EVIDENCIAS DE APRENDIZAJE.

Este subproceso tiene por objeto regular la recopilación y registro de las pruebas y evidencias que habrán de posibilitar la valoración del aprendizaje de acuerdo a lo establecido en la planificación. Centrado fundamentalmente en la fase de implementación de la evaluación, conlleva la consideración técnica consciente de los siguientes elementos (criterios).

3.1. Implementación ajustada.

Este primer criterio determina que el proceso de recopilación de las evidencias de aprendizaje se desarrolle de acuerdo a lo establecido en la fase previa de planificación. El cumplimiento coherente de la previsión evaluativa favorece la transparencia informativa ante las/os estudiantes y evita la incursión en sesgos o desviaciones que puedan comprometer la validez y fiabilidad del proceso. Requiere el equilibrio entre el desarrollo fiel de lo previsto en la planificación y la introducción de ajustes oportunos y justificados en respuesta a los imperativos e imprevistos prácticos.

El criterio se concreta en los dos siguientes rasgos evaluativos o indicadores, entre otros posibles:

3.1.1. Cumplimiento coherente de la planificación (consultar también 4.2.1.).

La puesta en práctica del sistema evaluativo y la recogida de evidencias se desarrolla en orden a los compromisos y decisiones adoptadas en la fase de planificación.

3.1.2. Gestión de los ajustes realizados.

Los ajustes o adaptaciones que exijan los imprevistos de la puesta en práctica, se analizan, documentan y resuelven de forma justificada, ofreciendo información oportuna a las personas implicadas y/o consensuando con ellas alternativas de solución.

3.2. La recopilación de evidencias da lugar a un registro personalizado sobre el aprendizaje y la evolución de cada estudiante, que permite el seguimiento individual de cada una/o de las/os integrantes del grupo a corto y largo plazo.

Este criterio delimita la creación de un sistema de registro personal de evolución y aprendizaje que, garantizando los derechos básicos de protección de la identidad del alumnado, permita efectuar un seguimiento individual, favoreciendo su mejor conocimiento y orientación. Dependiendo de la situación particular, estos registros pueden recabar evidencias de aprendizaje a corto o largo plazo y en formatos disciplinarios o interdisciplinarios.

El criterio se concreta en los dos siguientes rasgos evaluativos o indicadores, entre otros posibles:

3.2.1. Orientación y seguimiento individual (consultar también 1.4.2.).

Las evidencias recabadas a lo largo del proceso se incluyen en un registro individual de evolución y aprendizaje que permite efectuar el seguimiento de cada uno de los integrantes del grupo y adoptar las medidas particulares de orientación que se estimen oportunas.

3.2.2. Legalidad (consultar también 5.1.1. y 6.1.3.).

Los registros de seguimiento individual de cada estudiante se emplean única y exclusivamente con fines académicos y formativos, velando por los derechos básicos de protección de la identidad del alumnado.

3.4 Subproceso 04: análisis y calificación de resultados.

Definido a continuación, el cuarto subproceso se concreta en dos criterios y cinco indicadores.

4. ANÁLISIS Y CALIFICACIÓN DE RESULTADOS.

El cuarto subproceso centra la atención en el análisis que tiene lugar a partir de las evidencias de aprendizaje recopiladas y el proceso de calificación en el que éste desemboca en caso de evaluación acreditativa. Basado en la racionalidad y transparencia técnica del análisis y sus resultados, este nuevo subproceso pone especial atención en el desenlace de la finalidad sumativa de la evaluación y conlleva la consideración técnica consciente de los siguientes elementos (criterios).

4.1. El dispositivo evaluativo incorpora medidas para reducir los principales sesgos y desviaciones que pueden tener lugar durante el proceso de análisis de las evidencias y los resultados.

Este primer criterio atiende a la neutralización deliberada de los riesgos que pueden contaminar el proceso de análisis de los resultados y evidencias de aprendizaje y las decisiones que a partir del mismo se adopten. Se orienta fundamentalmente a preservar las dos grandes propiedades psicométricas sujetas a los procesos generales de análisis (la validez y la fiabilidad), planteando un ejercicio de cautela respecto a los mismos.

El criterio se concreta en los dos siguientes rasgos evaluativos o indicadores, entre otros posibles:

4.1.1. Neutralización de las amenazas frecuentes a la fiabilidad (consultar también 6.1.2.).

En el proceso de análisis de evidencias de aprendizaje se introducen medidas o mecanismos para contrarrestar deliberadamente las amenazas frecuentes a su precisión (correcciones a ciegas para neutralizar los sesgos de expectativa, controles aleatorios de segunda valoración, desorden estratégico, etc.).

4.1.2. Neutralización de las amenazas frecuentes a la validez (consultar también 6.1.2.).

En el proceso de análisis de evidencias de aprendizaje se introducen deliberadamente medidas o mecanismos para evitar su desviación con respecto a los aprendizajes que pretenden valorarse (plantillas de explicitación de criterios, tablas de cruce entre objetivos y criterios de evaluación, etc.).

4.2. El análisis de las evidencias recopiladas conduce a la realización de juicios o valoraciones de acuerdo al sistema de calificación establecido.

El segundo criterio se dedica expresamente al proceso de calificación, donde el análisis de las evidencias registradas a lo largo del recorrido formativo se traduce en una valoración acreditativa del aprendizaje logrado. Dicha calificación debe ser coherente con los criterios de referencia, fiable y racional y transparente.

El criterio se concreta en los tres siguientes rasgos evaluativos o indicadores, entre otros posibles:

4.2.1. Calificación coherente (consultar también 6.1.2.).

El proceso de calificación se realiza de acuerdo a los criterios explicitados en la fase de diseño y comunicados al comienzo de la implementación y/o durante la misma.

4.2.2. Calificación fiable (consultar también 6.1.2.).

La calificación discrimina de forma precisa distintos niveles de logro en los aprendizajes previstos, según el cumplimiento evidenciado en los criterios de establecidos a tal efecto.

4.2.3. Racionalidad y transparencia.

La calificación de cada estudiante está justificada y argumentada de forma precisa a partir de criterios explícitos y públicos.

3.5 Subproceso 05: feedback y proacción.

Definido a continuación, el quinto subproceso incorpora cuatro criterios y un total de 13 indicadores para su desarrollo.

5. FEEDBACK Y PROACCIÓN.

Este subproceso está centrado en a la comunicación de los resultados de aprendizaje revelados por medio de la evaluación en las distintas fases del proceso formativo. Comprende la comunicación formal y legal de los resultados constatados, así como los procesos de retroalimentación y proacción desarrollados a partir de los mismos para la mejora de la enseñanza y del aprendizaje. Su desarrollo debe efectuarse desde la consideración técnica consciente de los siguientes elementos (criterios).

5.1. El dispositivo evaluativo establece un proceso de comunicación de resultados que satisface las necesidades de las diversas audiencias interesadas y garantiza el cumplimiento de sus derechos.

El primer criterio hace referencia a la comunicación formal de los resultados constatados en la evaluación del aprendizaje, que debe desarrollarse respetando los derechos fundamentales de las audiencias involucradas (fundamentalmente los del alumnado) y en coherencia con lo establecido en la normativa evaluativa institucional.

Se concreta en los dos siguientes rasgos evaluativos o indicadores, entre otros posibles:

5.1.1. Informes legales (consultar también 6.1.3.).

Los resultados de aprendizaje se comunican en formatos que preservan los derechos fundamentales reconocidos a las/os estudiantes (anonimato, información, etc.).

5.1.2. Cumplimiento de la normativa institucional (consultar también 6.1.3.).

La comunicación de los resultados de aprendizaje tiene lugar de acuerdo a lo establecido en la normativa evaluativa de referencia (plazos, formatos, derechos y deberes, etc.).

5.2. El dispositivo evaluativo no se limita a juzgar los resultados constatados, sino que ofrece retroalimentación para contribuir a su comprensión profunda.

Este nuevo criterio atiende al feedback o retroalimentación que se establece a partir de la comunicación de resultados de aprendizaje a lo largo del proceso formativo. Más allá de la mera información, la comisión de feedback en determinadas condiciones, contribuye a comprender en profundidad el aprendizaje esperado y la situación actual con respecto al mismo. Ofrecido de forma frecuente, puntual y suficiente, sin que represente una amenaza para el alumnado y en referencia a los criterios de evaluación establecidos, constituye un apoyo fundamental para el aprendizaje y su mejora.

El criterio se concreta en los ocho siguientes rasgos evaluativos o indicadores, entre otros posibles:

5.2.1. Feedback referido a criterios y objetivos.

El feedback o retroinformación ofrecida al alumnado está exenta de valoraciones personales y centra su atención en la tarea y los criterios de evaluación y objetivos definidos.

5.2.2. Feedback puntual.

El feedback o retroinformación se ofrece de forma puntual, en un intervalo de tiempo en que aún tiene sentido para la/el estudiante y utilidad pedagógica.

5.2.3. Feedback frecuente.

El alumnado recibe feedback o retroinformación de forma frecuente y en pequeñas porciones, de modo que se favorezca su asimilación.

5.2.4. Feedback descriptivo, comprensible y suficiente.

El feedback o retroinformación incorpora la información y el nivel de detalle suficiente como para resultar comprensible a sus destinatarias/os.

5.2.5. Feedback dialógico.

El feedback se desarrolla en un proceso dialógico e interactivo en el que el alumnado tiene la oportunidad de aclarar sus dudas e inquietudes ante las expectativas de aprendizaje establecidas y su situación respecto a las mismas.

5.2.6. Feedback sostenible (consultar también 7.1.3.).

El feedback es generado y administrado por medio de procedimientos realistas y factibles, que integran la participación de diversos agentes para favorecer su viabilidad.

5.2.7. Feedback estimulante.

El feedback o retroinformación ofrecida al alumnado es segura, no representa una amenaza para su autoestima ni le genera daño, desmotivación o desánimo.

5.2.8. *Feedback prospectivo (consultar también 5.3.1.).*

El feedback debe incorporar orientación que ayude a la/al estudiante a comprender qué y cómo debe actuar en el futuro para mejorar el logro evidenciado.

5.3. El feedback proporcionado a las/os estudiantes incorpora pautas claras de proacción o mejora para reducir la distancia entre los resultados de aprendizaje constatados y los esperados.

Este tercer criterio hace referencia a la proacción ejercida sobre el aprendizaje, esto es, a la formulación de sugerencias y pautas de mejora mediante el feedback, para orientar el trabajo del alumnado en línea con las metas previstas.

El criterio se concreta en los dos siguientes rasgos evaluativos o indicadores, entre otros posibles:

5.3.1. *Claridad y utilidad de las consignas.*

La evaluación da lugar a pautas claras de revisión, actuación y mejora para favorecer en quien las recibe mayores cotas de calidad en próximas realizaciones.

5.3.2. *Seguimiento a partir del feedback.*

El dispositivo evaluativo establece cauces para supervisar la aplicación que hace el alumnado de las pautas de revisión, actuación y mejora desprendidas del feedback.

5.4. Los resultados de aprendizaje discente constatados en la evaluación se emplean para revisar, adecuar y mejorar el proceso de enseñanza.

Este último criterio se centra en la regulación de la enseñanza a partir de los resultados de aprendizaje observados en el alumnado. Dichos resultados suscitan un proceso de reflexión y revisión de la docencia que desemboca en la introducción de medidas para favorecer mejor el aprendizaje.

El criterio se concreta en el siguiente rasgo evaluativo o indicador, entre otros posibles:

5.4.1. *Mejora de la docencia.*

Se introducen ajustes y adecuaciones en el proceso de enseñanza en función de los resultados de aprendizaje discente evidenciados en la evaluación.

3.6 Subproceso 06: revisión del proceso de evaluación (metaevaluación).

El sexto subproceso queda conformado por dos criterios y seis indicadores:

6. REVISIÓN DEL PROCESO DE EVALUACIÓN (META-EVALUACIÓN).

El sexto subproceso es la metaevaluación, que plantea la revisión del dispositivo evaluativo diseñado e implementado en la valoración del aprendizaje discente. Con intervención de uno o más agentes de evaluación, comprende el análisis sistemático del proceso de evaluación, al menos en sus dimensiones ética, técnica y pedagógica, para identificar sus principales limitaciones e introducir mejoras. Ello exige la consideración técnica consciente de los siguientes elementos (criterios).

6.1. Se establece un proceso de revisión específica sobre el sistema evaluativo empleado para la valoración del aprendizaje discente.

Este primer criterio hace referencia a la metaevaluación o revisión del sistema evaluativo. Plantea la revisión de las posibilidades y limitaciones técnicas, éticas y pedagógicas del dispositivo evaluativo diseñado y puesto en práctica para producir información que contribuya a su mejora.

El criterio se concreta en los cuatro siguientes rasgos evaluativos o indicadores, entre otros posibles:

6.1.1. Valoración pedagógica.

Se realiza una revisión sistemática de las posibilidades y limitaciones del dispositivo evaluativo como herramienta didáctica para la promoción del aprendizaje del alumnado (¿representa un apoyo útil?, ¿favorece una aproximación adecuada al aprendizaje?, etc.).

6.1.2. Valoración técnica.

Se realiza una revisión sistemática de las posibilidades y limitaciones del dispositivo evaluativo configurado, como herramienta técnica para la acreditación de los aprendizajes del alumnado (¿es válida?, ¿es fiable?, ¿resulta viable?, ¿es sistemática?, etc.).

6.1.3. Valoración ética.

Se realiza una revisión sistemática de las posibilidades y limitaciones de la evaluación para favorecer el cumplimiento efectivo de los derechos y principios básicos asociados a su puesta en práctica (justicia, transparencia, beneficencia, privacidad, etc., según la política evaluativa de referencia).

6.1.4. Metaevaluación participativa.

En la revisión de los dispositivos evaluativos participan el profesorado responsable y otros agentes evaluadores (proferentemente, el alumnado).

6.2. Los resultados de la metaevaluación se emplean para mejorar la calidad ética, técnica y/o pedagógica de los dispositivos evaluativos empleados en la evaluación del aprendizaje discente.

Este segundo criterio es consecuencia del anterior y hace referencia a la gestión de la información producida por la metaevaluación. Dicha revisión debería contribuir a una identificación realista de las limitaciones y posibilidades del dispositivo evaluativo articulado, así como a la introducción de ajustes y mejoras sobre el mismo.

El criterio se concreta en los dos siguientes rasgos evaluativos o indicadores, entre otros posibles:

6.2.1. Asunción de posibilidades y limitaciones.

Las posibilidades y limitaciones de la evaluación como instrumento para acreditar y promover el aprendizaje están claramente identificadas y documentadas.

6.2.2. Mejora del dispositivo evaluativo.

Se introducen ajustes y adecuaciones en el sistema evaluativo en función de los resultados y conclusiones reveladas en el proceso de metaevaluación.

3.7 Subproceso 07: significación didáctico-pedagógica de la experiencia evaluativa.

Definido a continuación, el último subproceso se concreta en un único criterio y tres indicadores articulados para su desarrollo.

7. SIGNIFICACIÓN DIDÁCTICO-PEDAGÓGICA DE LA EXPERIENCIA EVALUATIVA.

El séptimo subproceso se centra en el especial papel que juega la evaluación del aprendizaje en el contexto específico de la formación básica de maestras/os. Además de representar una herramienta didáctica y acreditativa, en estas titulaciones la evaluación es una de las competencias a desarrollar por el alumnado. En este contexto el sistema evaluativo empleado por el profesorado, constituye en sí mismo una experiencia de aprendizaje. En consecuencia, lejos de replicar las disfuncionalidades típicas, los sistemas evaluativos deberían estar conscientemente ideados para favorecer concepciones evaluativas deseables entre las/os futuras/os evaluadoras/es. Para ello, se precisa la consideración técnica consciente del siguiente criterio:

7.1. El dispositivo evaluativo prevé deliberadamente la creación de experiencias significativas de modelado para el desarrollo de la competencia evaluativa de las/os futuras/os maestras/os.

El criterio se centra específicamente en la doble significación de la evaluación del aprendizaje en la formación básica de maestras/os, donde además de constituir un proceso educativo clave, representa una de las competencias específicas cuya consecución se pretende. Esta especial circunstancia debe conducir a la configuración deliberada de experiencias evaluativas ejemplares que modelen, desde la vivencia directa, representaciones y prácticas evaluativas en línea con los supuestos teóricos.

El criterio se concreta en los tres siguientes rasgos evaluativos o indicadores, entre otros posibles:

7.1.1. Evaluación coherente.

El dispositivo evaluativo diseñado integra de forma consistente los principios que orientan la formación de las/os futuras/os maestras/os en materia evaluativa, evitando deliberadamente la incursión en incoherencias que refuercen las concepciones evaluativas clásicas y resten credibilidad al discurso teórico.

7.1.2. Evaluación ejemplar.

El sistema de evaluación diseñado modela los principios teóricos que deben inspirar la evaluación del aprendizaje, ofreciendo un ejemplo real de buena praxis para los futuras/os maestras/os y evaluadoras/es.

7.1.3. Posicionamiento docente.

El sistema de evaluación incorpora deliberadamente actividades de evaluación participativa, en las que las/os futuras/os maestras/os experimenten el rol evaluador que habrán de desempeñar en el futuro.

4. El capítulo a grandes rasgos

El capítulo que concluye ha explicitado la particular visión o teoría de la evaluación del aprendizaje sobre la que se asienta la presente investigación. El resultado es el modelo ITC (Integrado Teórico Complejo) que precisa siete subprocesos en la evaluación del aprendizaje universitario y un sistema de criterios e indicadores para el desarrollo de cada uno de los mismos. El modelo está ideado para el contexto específico de la formación básica de maestras/os de Educación Infantil y Primaria, aunque es fácilmente extrapolable a otras titulaciones suprimiendo el último subproceso y adecuando la terminología que emplea.

La propuesta del modelo ITC ha sido definida y justificada partiendo del análisis de tres tipos referentes relevantes:

- La *conceptualización teórica de la evaluación para el aprendizaje* representa el primer pilar en el que se apoya el modelo ITC. De la misma incorpora toda una serie de rasgos y orientaciones para hacer de la evaluación un recurso para la estimulación consciente y deliberada del aprendizaje discente, sin limitarse a su acreditación.
- El *contexto específico* en el que se pretende aplicar el modelo ITC de evaluación representa la segunda referencia considerada en su configuración. Este contexto está determinado por tres factores relevantes en el momento de su definición:
 - ✓ Las implicaciones evaluativas que supone la consolidación del paradigma formativo asociado al EEES.
 - ✓ Los principales referentes normativos que condicionan el desarrollo de las prácticas evaluativas a nivel nacional y europeo.

- ✓ La especial significación que adopta la evaluación del aprendizaje en la formación básica de maestras/os, profesión regulada en la que conforma uno de los grandes aprendizajes a realizar por el alumnado para su futuro ejercicio docente.
- La *revisión de aquellos antecedentes y propuestas previas* que han tratado de sistematizar la evaluación del aprendizaje en la formación básica de maestras/os representa la tercera fuente inspiradora. Para definir los subprocesos, criterios e indicadores recogidos en el modelo ITC se han considerado un total de 15 antecedentes con desigual nivel de sistematización (10 decálogos o manifiestos de reconocido prestigio, tres propuestas estructuradas para la aplicación institucional, y dos propuestas de estándares de evaluación del aprendizaje de las/os estudiantes).

El resultado es un conjunto ordenado de 19 criterios y 68 indicadores que desglosan lo que en la presente investigación se considera una práctica evaluativa de calidad (**ANEXO 04**). La propuesta de competencia evaluativa desarrollada en el próximo capítulo parte de esta particular visión y trata de orientar la actuación docente hacia ella.

CAPÍTULO 5.

COMPETENCIA EVALUATIVA DEL PROFESORADO UNIVERSITARIO DE LOS GRADOS DE EDUCACIÓN INFANTIL Y PRIMARIA: PROPUESTA DE UN MODELO ESPECÍFICO PARA EL DIAGNÓSTICO Y LA FORMACIÓN

El capítulo quinto del marco teórico centra su atención en la competencia evaluativa del profesorado vinculado a la formación básica de maestras/os, núcleo central del estudio empírico planteado. Describe la construcción de un modelo teórico para encauzar el análisis y la formación de la misma, como fruto de una revisión y reflexión más amplia de la que da cuenta la figura 20:

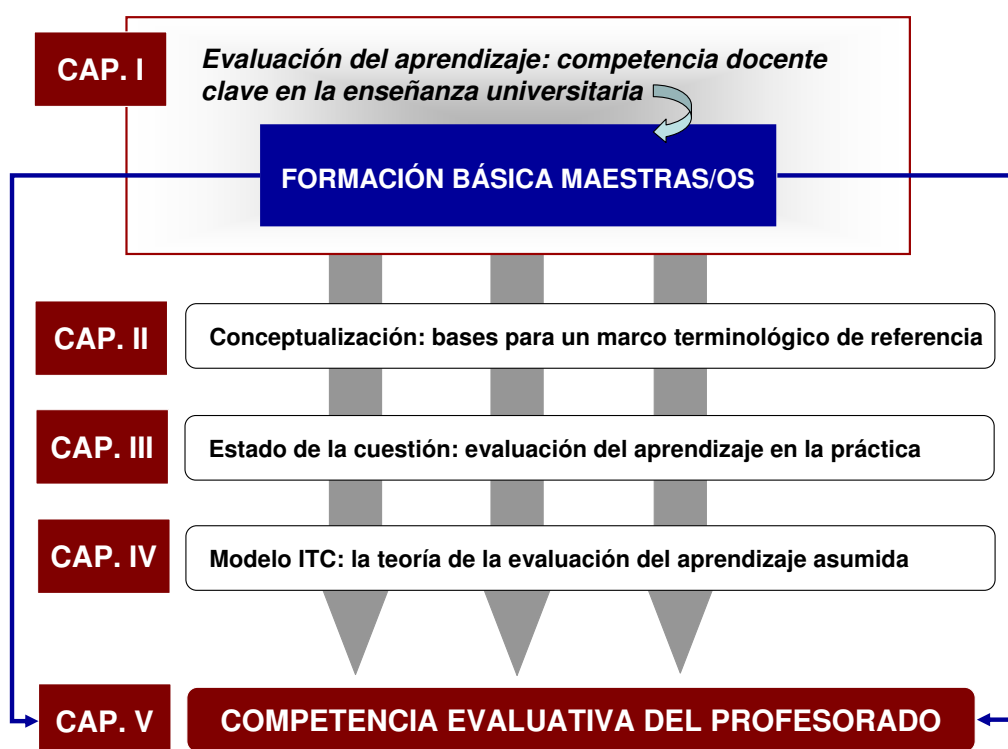


Figura 20. Competencia evaluativa del profesorado: conclusión del marco teórico.

Realizadas las acotaciones y delimitaciones conceptuales básicas en los dos primeros capítulos, el recorrido dibujado hasta el momento ha propiciado la respuesta

tentativa a dos preguntas enfrentadas: cómo tiene lugar la evaluación del aprendizaje en la práctica (el estado de la cuestión) y cómo debería llevarse a cabo desde la perspectiva teórica (modelo Integrado Teórico Complejo- ITC). El avance siguiente en esta secuencia de interrogantes, conduce a la delimitación de la competencia evaluativa del profesorado universitario, de modo que se establezca cuáles son los indicadores a los que se debe ajustar la actuación docente para favorecer la producción de prácticas evaluativas de calidad, en línea con lo dispuesto en el modelo ITC.

Las próximas páginas contienen la propuesta de modelo de competencia evaluativa elaborado en respuesta a esta última pregunta. Para su presentación, el capítulo se divide en cuatro grandes apartados, a los que se suma una síntesis final. Los tres primeros realizan la descripción general de la génesis de la propuesta, mientras que el cuarto presenta directamente el modelo resultante:

- El primer apartado enumera los rasgos generales que caracterizan el modelo de competencia evaluativa docente elaborado y las dos grandes utilidades para las que ha sido concebido.
- El segundo apartado detalla el proceso recorrido en su definición, indicando brevemente el trabajo realizado en cada paso.
- El tercer apartado explicita los cuatro tipos de fuente o referente sobre los que se fundamenta la propuesta de sistematización elaborada.
- El apartado cuarto presenta el modelo de competencia evaluativa resultante, recorriendo cada una de las ocho dimensiones que incorpora, así como los indicadores y descriptores definidos para su desarrollo.

1. Competencia evaluativa del profesorado universitario dedicado a la formación básica de maestras/os: presentación del modelo

La competencia evaluativa del profesorado ha sido definida como la respuesta eficaz y responsable a los dilemas y demandas de la evaluación del aprendizaje discente en el contexto formativo de referencia, en este caso, la formación básica de maestras/os. El modelo presentado en las próximas páginas identifica ocho dimensiones interrelacionadas

que intervienen en la misma (ver tabla 44); juntas ofrecen la imagen de la capacitación general que precisa el profesorado para realizar evaluaciones de calidad acordes al modelo ITC.

Tabla 44

Dimensiones de la competencia evaluativa del profesorado universitario en los grados dedicados a la formación básica de maestras/os de Educación Infantil y Primaria

1. DIMENSIÓN	Planificación estratégica de la evaluación del aprendizaje. <i>Planificar estratégicamente la evaluación del aprendizaje para promover en las/os estudiantes el logro de las competencias previstas en el currículo y su acreditación, a partir de una adecuada cantidad y variedad de evidencias recabadas a lo largo de todo el proceso formativo.</i>
2. DIMENSIÓN	Clarificación de criterios y condiciones. <i>Clarificar con las/os estudiantes los criterios y condiciones que rigen la evaluación del aprendizaje para establecer una visión clara y unívoca de la demostración de los logros a alcanzar que oriente su trabajo a lo largo del proceso.</i>
3. DIMENSIÓN	Atención a la diversidad en la evaluación del aprendizaje. <i>Adaptar las condiciones sobre las que se realiza la evaluación del aprendizaje para responder a la diversidad de necesidades de las/os estudiantes, considerando los criterios mínimos definidos de forma general y preservando la transparencia y justicia del proceso evaluativo.</i>
4. DIMENSIÓN	Implementación del sistema de evaluación. <i>Implementar los sistemas de evaluación planificados para garantizar el cumplimiento de sus propósitos, introduciendo y documentando los ajustes necesarios para responder a las condiciones prácticas reales y a los imprevistos surgidos.</i>
5. DIMENSIÓN	Análisis e interpretación de las evidencias de aprendizaje. <i>Analizar e interpretar las evidencias de aprendizaje recopiladas para generar valoraciones válidas y precisas del proceso de aprendizaje y sus resultados, aplicando medidas que contrarresten las principales causas de sesgo e influencia.</i>
6. DIMENSIÓN	Regulación del aprendizaje y de la enseñanza. <i>Utilizar los resultados constatados en el proceso de evaluación para regular el aprendizaje y la enseñanza, identificando logros y áreas de mejora y estableciendo pautas de actuación oportunas para su superación.</i>
7. DIMENSIÓN	Participación discente en la evaluación del aprendizaje. <i>Utilizar la participación activa de las/os estudiantes en el proceso evaluativo para contribuir a la autorregulación progresiva de su aprendizaje hacia las metas formativas definidas.</i>
8. DIMENSIÓN	Revisión del sistema de evaluación. <i>Revisar el sistema de evaluación para identificar fortalezas y debilidades que puedan orientar su mejora desde la perspectiva ética, técnica y pedagógica.</i>

El modelo configurado representa una visión sintética o teoría pedagógica (B. García et al., 2008) sobre la competencia del profesorado en la evaluación de los aprendizajes de

las/os estudiantes. En consecuencia, plantea una imagen parcial de este relevante espacio de la actividad docente, aunque bien fundamentada. Las fuentes sobre las que se justifica su construcción, así como el proceso recorrido en su definición se presentan en próximos apartados.

La propuesta de sistematización de la competencia evaluativa se realiza desde un enfoque multidimensional y complejo, en respuesta a la propia naturaleza del proceso evaluativo, que así lo exige. En concreto, la estructura del modelo se desarrolla en tres niveles progresivos de definición que explicitan una determinada manera de ejercer la competencia evaluativa.

- El primer nivel de análisis determina que la competencia evaluativa se encuentra integrada por *ocho dimensiones* involucradas en la evaluación del aprendizaje (tabla 44). Las dimensiones representan las distintas facetas o aspectos que pueden considerarse en un asunto (Real Academia Española, 2014), en nuestro caso, en la evaluación del aprendizaje. Estrechamente interrelacionadas e interdependientes, la identificación de estas dimensiones clave responde a un artificial ejercicio de delimitación, efectuado para facilitar un análisis más sistemático y pormenorizado. Todas ellas responden a un mismo proceso y, en consecuencia, no mantienen entre sí límites claros ni pueden desligarse.

En definitiva, el establecimiento de estas grandes dimensiones dentro de la competencia evaluativa, responde a la siguiente pregunta clave: ¿qué debe ser capaz de hacer y demostrar el profesorado para consolidar prácticas evaluativas de calidad en los términos definidos en el modelo ITC?

- El segundo nivel concreta *indicadores* por cada uno de los aspectos o dimensiones considerados en la evaluación del aprendizaje. Estos indicadores establecen una selección de las actuaciones evaluativas que se consideran necesarias en cada dimensión, por comprometer la calidad de la evaluación del aprendizaje que define el modelo ITC. “Un indicador no es más que una señal que permite captar y representar aspectos de una realidad que no son directamente accesibles al observador” (Tejedor, 2003, p. 161). La pregunta clave que guía su delimitación sería la siguiente: ¿qué actuaciones concretas son necesarias para lograr el reto expresado en cada una de las dimensiones evaluativas?
- El tercer nivel detalla *descriptores de logro* para cada uno de los indicadores formulados, delimitando diferentes niveles de consecución en los mismos. El modelo presentado incorpora descriptores para cinco posibles niveles de logro,

ordenados de menor a mayor grado de cumplimiento: insuficiente o inadecuado, poco adecuado, adecuado, muy adecuado y excelente. Los descriptores resumen el tipo de manifestaciones prácticas previsibles en cada nivel, por lo que la pregunta que lleva a su formulación se adapta en cada caso: ¿qué tipo de rasgos y conductas se hacen evidentes cuando el cumplimiento del indicador es inadecuado/ poco adecuado/ adecuado/ muy adecuado o excelente?

En suma, los tres niveles de desarrollo recorridos describen la competencia evaluativa del profesorado y la operativizan en una rúbrica o rejilla de valoración expuesta en el **ANEXO 05**. Siguiendo las orientaciones de Blanco (2008b) ésta pueden clasificarse de acuerdo a dos criterios:

- Desde el punto de vista temático es de carácter específico, en la medida en que responde a una tarea particularmente ligada al ejercicio de la docencia (la evaluación del aprendizaje).
- Atendiendo a su estructura o concepción formal es analítica, ya que cada uno de los indicadores considerados se valora por separado y en una escala descriptiva propia⁵⁹.

Presentado en la parte final del capítulo, sobre el modelo resultante cabe realizar ciertas aclaraciones de partida:

- La propuesta de sistematización presentada no debe interpretarse como un esquema cerrado de actuaciones que pueda ser entrenado de forma fragmentada por el profesorado. La competencia evaluativa exige la combinación armónica de recursos de diversa naturaleza en situaciones variables, de modo que debe manejarse como una referencia abierta y flexible. Si en el capítulo primero se proponía para el profesorado un perfil docente transferencial, flexible y polivalente (Bozu & Canto, 2009; Valcárcel, 2003), tales rasgos también son aplicables a la delimitación de esta competencia. Así pues, no constituye un instrumento para la unificación estricta de las prácticas evaluativas; puede adecuarse a estilos docentes muy diversos, responder a la especificidad de áreas de conocimiento diferentes y dar lugar a grados desiguales de desarrollo.

⁵⁹ Cuando las rúbricas consideran la valoración de productos totales y de forma global, reciben la denominación de holísticas (Blanco, 2008b; De la Orden, 2011b), comprensivas o totales (Mertler, 2001). Desde el punto de vista temático, las rúbricas que detallan competencias generales o transversales se les denomina genéricas.

- Aunque aporta orientaciones teóricas, el modelo propuesto está ideado para ser abordado desde la práctica. Como asevera Mateo (2012), “no hay posibilidad de enseñar a nadie a ser competente, esta es una característica que únicamente se desarrolla y se obtiene a partir del aprendizaje sobre actividades realizadas en diversos contextos [...]” (p. 219).
- En virtud del carácter reconstructivo que acompaña a toda competencia (Cano, 2007, 2011; Le Boterf, 2001; Pavié, 2011), el modelo ideado ofrece un marco de referencia para el desarrollo profesional a largo plazo. No debe ser considerado como una enumeración de mínimos o prerrequisitos, sino como ideal hacia el que aproximar progresivamente la práctica evaluativa.
- Esta propuesta de sistematización responde específicamente al perfil docente del profesorado universitario dedicado a la formación básica de maestras/os. Contextualizada en los Grados de Educación Infantil y Primaria, su formulación se propone suponiendo una cierta cualificación inicial del profesorado destinatario que permita su seguimiento. Sin embargo, podría ser fácilmente adaptada para orientar la revisión y formación de la competencia evaluativa de docente en otras titulaciones.

Partiendo de estas premisas, la propuesta de competencia evaluativa ha sido ideada para servir a dos grandes utilidades interrelacionadas: la evaluación y la formación. Presentadas en los próximos apartados, su desarrollo práctico se articula presumiendo la confluencia de unas condiciones de partida:

- Tanto en la modalidad formativa, como en la evaluativa, la aplicación se plantea desde una *orientación exclusivamente pedagógica*, no para respaldar procesos de acreditación. En línea con las recomendaciones de B. García et al. (2008) o Tejedor (2003), el modelo está concebido para permitir al profesorado retroalimentar la docencia y revisarla para su mejora.
- Además de pedagógica, la aplicación del modelo está pensada para ser *voluntaria*. Esta segunda condición responde a un proceso de deliberación en el que se han confrontado argumentos procedentes de diversas posturas. De un lado, se sitúan quienes sostienen que la profesionalización de la docencia exige alejarse del voluntarismo (Michavila, 2005; Santos, 2010a) y del otro, quienes alertan de los riesgos de incurrir en procesos de formación o evaluación reducidos, que no llegan a extenderse ni a generalizarse por su escasa significación (Ahumada, 2002; Poblete, 2007). Admitidas ambas visiones, consideramos que mejorar la competencia

evaluativa, exige adoptar una actitud crítica de revisión (Bilbao, 2010; Race, 2003) y el reconocimiento de la escasa autenticidad de muchas de las prácticas evaluativas desarrolladas (Ahumada, 2005). Esta condición es sentida y difícilmente puede forzarse, de modo que la propuesta plantea restringir la participación a aquellas/os docentes que sientan la necesidad de hacer formación y/o revisión en materia evaluativa. Así lo resuelven también Margalef y Álvarez (2005) y Parcerisa (2010).

- Según se detalla en cada modalidad, el modelo presentado puede ser aplicado consensuando distintos niveles de intervención e implicación entre el profesorado:
 - ✓ En cuanto a la responsabilidad de la dirección, la evaluación y/o formación de la competencia evaluativa docente puede desarrollarse contando con la colaboración de agentes externos o sin ella (proceso de revisión o formación autodirigida). Esta segunda vía autodirigida representa uno de los cauces más factibles para el desarrollo profesional del profesorado a juicio de A. Medina et al. (2011) o Caballero (2013).
 - ✓ Asimismo, las iniciativas de evaluación y formación pueden desarrollarse a nivel individual (un/a docente revisando su competencia evaluativa) o en equipo (varias/os profesoras/es realizando dicho ejercicio de forma simultánea y más o menos compartida). Esta última alternativa encuentra un respaldo creciente en la literatura y la investigación especializada porque suscita procesos de aprendizaje lateral (Delbanco, 2012), diálogos e intercambios enriquecedores y la creación de redes de coordinación y apoyo (Blanch & Corcelles, 2014; Mauri et al., 2013; Prebble et al., 2004; Southwell & Morgan, 2009).
- Las dos grandes funcionalidades atribuidas al modelo no son incompatibles. Formación y evaluación deberían tener lugar de forma conjunta y retroalimentarse, dando lugar a distintas secuencias de aplicación; dependiendo de las circunstancias particulares en las que vaya a tener lugar la aplicación y de las prioridades que la motiven, cabe la posibilidad de comenzar realizando un diagnóstico de la competencia evaluativa para establecer áreas de formación en consecuencia o a la inversa, empezar trabajando sobre la propuesta como referencia formativa, y acometer más tarde su evaluación específica (antes de desarrollar una acción concreta de mejora sobre el sistema evaluativo y/o una vez finalizada, por ejemplo).

1.1 Evaluación a partir del modelo de competencia evaluativa propuesto.

El modelo de competencia evaluativa propuesto puede ser empleado como marco de referencia para el desarrollo de procesos de revisión o metaevaluación sobre los sistemas de evaluación aplicados. Planteado a modo de rúbrica específica y analítica, la delimitación de dimensiones, indicadores y descriptores de logro representa un recurso operativo para valorar la manifestación práctica de la competencia evaluativa docente.

Esta primera utilidad remite a los beneficios generales declarados para la metaevaluación en el capítulo tercero y suma dos contribuciones adicionales específicamente relacionadas con la propuesta:

- Por una parte, contribuye a superar la indefinición de los aspectos a revisar, una de las mayores dificultades que acompaña a los procesos de metaevaluación (Schwandt & Halpern, 1988). Ante la confusión y la falta de acuerdo que rodea la praxis evaluativa, ofrece una referencia sobre la que apoyar la detección de fortalezas y áreas de mejora
- Por otra parte, plantea un análisis revisión que va más allá de la perspectiva institucional de la evaluación, habitualmente la más extendida en los procesos de revisión (así se constata, por ejemplo, en Astin et al., 1992; ENQA, 2009; QAA, 2011). En su lugar, se centra de forma específica en la evaluación del aprendizaje del alumnado, remitiendo a un contexto de aplicación más próximo y significativo: su manifestación en la interacción didáctica entre profesorado y estudiantes en la dinámica cotidiana. Además de aportar otro enfoque, esta perspectiva no es incompatible con orientaciones más amplias sobre metaevaluación, como las ofrecidas por Scriven (2011), Stufflebeam (1999) o la American Evaluation Association (AEA, 2004).

También enlaza con los beneficios asociados a la evaluación de la actividad docente. Consideramos que las ventajas generales que reporta Riegle (1987) para la revisión de la calidad de la enseñanza⁶⁰ se magnifican particularmente al centrar la atención en la competencia evaluativa:

⁶⁰ Riegle (1987) añade un quinto beneficio ligado a la promoción de la carrera docente y el desarrollo de la trayectoria profesional. No ha sido mencionado por incompatibilidad con la orientación declaradamente pedagógica y no acreditativa de la revisión propuesta.

- *Desarrollo técnico-pedagógico.* Por su proximidad con la planificación y la articulación metodológica, la revisión de la competencia evaluativa puede dar lugar al desarrollo de las habilidades didáctico-pedagógicas ligadas a la enseñanza.
- *Desarrollo profesional.* Destacada como una de las parcelas más complejas de la actividad docente, el análisis pormenorizado de la competencia evaluativa puede proporcionar una fuente destacada de formación y aprendizaje relevante, contribuyendo al crecimiento individual del profesorado en el contexto institucional y social de la profesión.
- *Desarrollo organizativo.* La revisión de la competencia evaluativa puede favorecer el ajuste y modificación de las condiciones estructurales que rodean la evaluación, un factor clave de la calidad institucional (Gibbs, 2010; Kaneko, 2008; Villa, 2004). Presentadas en el capítulo tercero, estas condiciones remiten a factores institucionales y organizativos que, a menudo, ejercen un influjo negativo en la actividad evaluadora.
- *Desarrollo personal.* Sin perjuicio de lo anterior, la revisión la competencia evaluativa puede generar el crecimiento personal del profesorado, al provocar la reflexión sobre las implicaciones y repercusiones éticas que la rodean (valores y principios de referencia, dilemas, etc.).

Anteriormente justificadas, las premisas de voluntariedad y finalidad pedagógica adoptan una especial significación en el empleo evaluativo del modelo, para evitar los celos y actitudes de autodefensa. Salvadas dichas condiciones, el proceso de metaevaluación puede recurrir a distintas fuentes o agentes informantes. Atendiendo a ese criterio, diferenciamos dos grandes modalidades de revisión, en las que cabe la dirección externa e interna y la intervención individual o grupal previamente referidas. Como en cualquier otra evaluación, el punto de partida es que cuantas más fuentes se consulten, mayor será la riqueza y credibilidad del diagnóstico resultante (B. McDonald, 2010):

- La *autoevaluación* recurre sólo a la perspectiva de quien ha diseñado y aplicado el dispositivo evaluativo y representa la opción más factible. Inherente a la enseñanza (B. García, Valencia, & Pineda, 2012), constituye una actividad basada en la capacidad del profesorado de poner en marcha habilidades metacognitivas para detectar sus fortalezas y debilidades y poder avanzar hacia un mejor desempeño (Fuentes & Herreros, 1999). Está sujeta al clásico sesgo de imprecisión evaluativa por deseabilidad social (Morales, 2006b), pero la participación voluntaria y la

finalidad pedagógica de la experiencia hace que los riesgos no sean severos en este caso.

- Junto a ella, pueden desarrollarse *modalidades híbridas o mixtas de evaluación* con participación de otros agentes informantes. La consulta puede ampliarse a colegas o superiores, al alumnado egresado y, principalmente a las/os estudiantes. Aunque sus valoraciones no produzcan cambio por sí mismas, tienen valor como fuente de información y es conveniente contemplarlas (Cisneros-Cohernour & Stake, 2010; Richardson, 2005; Williams & Brennan, 2003). En el contexto de una evaluación pedagógica y voluntaria, basada en la delimitación previa y explícita de una serie de dimensiones, indicadores y descriptores, la participación discente se desarrollaría en las condiciones recomendadas por Mateo (2000):

A nuestro juicio y a pesar del valor que todo el mundo adjudica a esta fuente de información, consideramos que la reflexión evaluativa de los alumnos respecto de sus profesores debe inscribirse en el marco exclusivo de la evaluación formativa del profesor y en su enfoque debe primarse el acentuar las posibilidades que encierra como instrumento para mejorar la relación alumno-profesor. Que el profesor conozca las percepciones del alumno sobre su práctica docente y su actuación personal en el aula, independientemente de su valor formativo, puede ser una plataforma de interacción y acercamiento entre los alumnos y el profesor de primerísima magnitud. (p. 21).

Independientemente del número de fuentes a las que se recurra y del nivel de participación que se acuerde en cada caso⁶¹, la revisión a partir del modelo de competencia presentado no debería plantearse como un fin en sí misma, sino como un proyecto más amplio de innovación. Tejedor (2003) lo expresa en los siguientes términos:

La evaluación no se limitará al papel de un mecanismo de optimización de recursos, lo cual es importante pero insuficiente, sino que ha de convertirse en un dispositivo de innovación, apuntando no sólo al análisis de fortalezas y debilidades (internas y externas) sino también a la reflexión sobre los objetivos y la manera de lograrlos; la

⁶¹ Con dirección interna o externa y de forma individual o grupal.

evaluación debe ser vista como una oportunidad para el debate del *deber ser*, para cuestionar los propósitos del quehacer universitario, sus modelos de actuación y su papel en el mundo actual. (p. 164).

Dicho proyecto enlaza de forma natural con la formación, segunda utilidad a la que pretende contribuir la propuesta de sistematización elaborada.

1.2 Formación a partir del modelo de competencia evaluativa propuesto.

Las dimensiones, indicadores y descriptores de logro definidos en la propuesta de sistematización (**ANEXO 05**) representan una hoja de ruta para guiar el diseño y la puesta en práctica de los procesos evaluativos, no sólo para su revisión. En ello estriba la funcionalidad formativa de la propuesta, que debería ir enlazada con la evaluación en los términos que ya han sido descritos.

Como dice J. M. Escudero (2006), “la formación docente debe entenderse en sentido amplio, abarcando tiempos, lugares y modalidades diversas” (p. 46). Desde esta perspectiva el modelo propuesto se concibe como un recurso flexible, aplicable a procesos de formación inicial y permanente, y abierto a modalidades diferentes de organización y participación⁶². Su éxito potencial en la promoción de la competencia evaluativa docente depende del uso concreto que se haga del mismo como herramienta formativa; no obstante, su delimitación inicial ha sido deliberadamente ideada para contribuir al cambio en las concepciones evaluativas profundas del profesorado, según lo expuesto en el capítulo tercero. En contraste con las orientaciones sugeridas para la formación evaluativa en el mismo, advertimos que la propuesta de sistematización elaborada se presta a las siguientes oportunidades:

- Plantea la praxis evaluativa como una realidad compleja (A. I. Pérez, 2010b; Santos, 2010a).
- Ofrece un recurso para que la formación se aproxime a las preguntas y problemas derivados de la práctica evaluativa cotidiana (Prebble et al., 2004; Southwell &

⁶² Como en la evaluación, puede ser liderada interna o externamente y desarrollada individualmente o en equipo.

Morgan, 2009). Puede ser puesto al servicio de acciones formativas contextualizadas o in situ (Santos, 2003) que hagan del trabajo cotidiano un espacio para el desarrollo profesional (Rodrigo, 1997).

- Estimula la reflexión de las/os participantes y favorece la integración de teoría y práctica desde una perspectiva global (Korthagen, 2010; M. Martínez, 2006; Zeichner, 2010).
- Propicia la participación directa del profesorado en la formación (Mateo, 2012; Mateos & Pérez, 2006). En aquellas modalidades de aplicación compartida, favorece además el intercambio y contraste de perspectivas (Pozo et al., 2006; Rodrigo, 1993).

2. Proceso de definición y ajuste

La delimitación de la competencia evaluativa docente ha tenido lugar a lo largo del proceso representado en la figura 21:

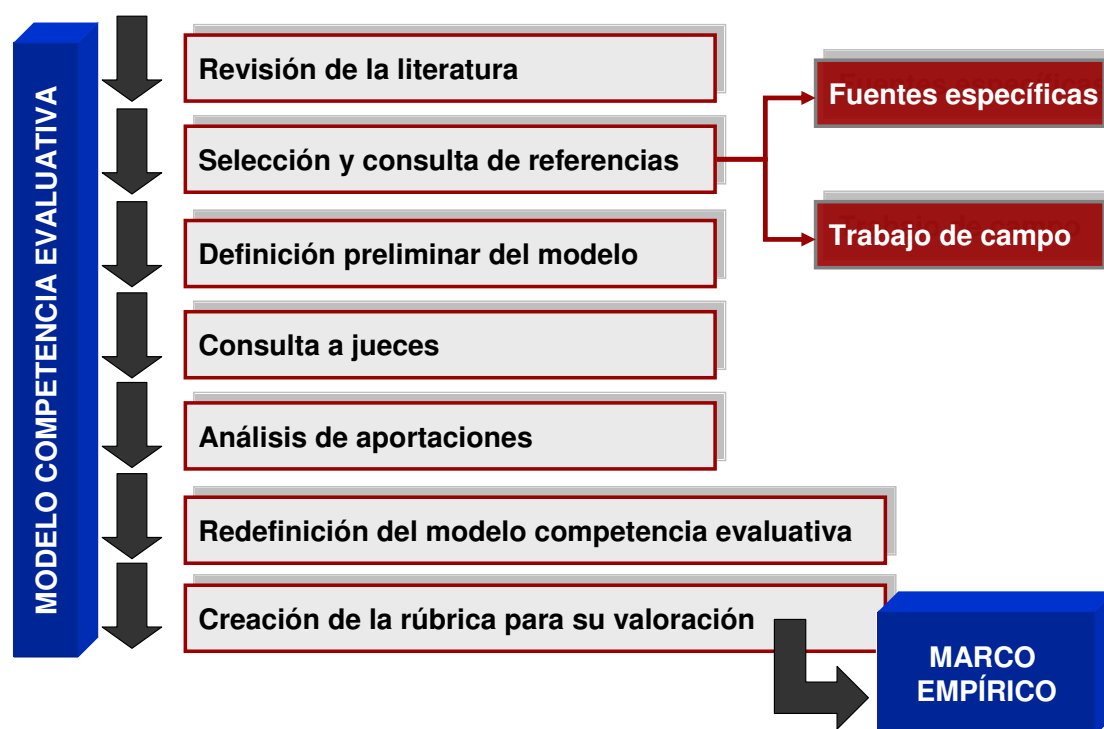


Figura 21. Proceso de definición del modelo de competencia evaluativa docente.

Su desarrollo ha supuesto el recorrido por siete fases consecutivas de trabajo que conectan el cuerpo teórico de la investigación con su desarrollo empírico:

- **Revisión de la literatura.** Esta primera fase ha llevado a definir de forma general la competencia evaluativa docente y a su contextualización preliminar en la formación básica de maestras/os.
- **Selección y consulta de referencias para la delimitación de la competencia evaluativa docente.** Partiendo de la revisión anterior, la segunda fase del proceso se ha centrado en la selección de referencias sobre las que construir el modelo de competencia evaluativa docente para las titulaciones reseñadas. Expuestas en el próximo apartado, ello ha supuesto el análisis de dos tipos de fuente:
 - ✓ *Revisión de fuentes específicas.* El punto de partida para la delimitación de la propuesta es la teoría de la evaluación explicitada en el modelo ITC. Junto a ella, se han considerado también las contribuciones de trabajos previos que se han aproximado a la sistematización de la competencia evaluativa docente, en la formación básica de maestras/os o en otros contextos formativos.
 - ✓ *Desarrollo del trabajo de campo sobre los tópicos evaluativos presentes en las escalas de evaluación de la calidad docente.* Como complemento de las fuentes específicas, se ha llevado a cabo el análisis de 63 escalas de evaluación de la calidad docente aplicadas en instituciones formativas que imparten los Grados de Educación Infantil y Primaria en todo el estado. El objetivo del estudio ha sido detectar qué aspectos de la evaluación del aprendizaje recogen dichas herramientas.
- **Definición preliminar de un modelo de competencia evaluativa.** El análisis de los referentes seleccionados ha conducido a una propuesta inicial de diez competencias evaluativas⁶³. Para refinar la propuesta, confirmar su pertinencia y solventar ciertas dudas internas, se ha desarrollado una cuarta fase de consulta a jueces.
- **Consulta a jueces.** En esta fase se ha solicitado colaboración a 32 expertas/os de reconocido prestigio en materia evaluativa, obteniendo respuesta de 18 de ellas/os (el 56,25%). La colaboración requerida ha girado en torno a tres propósitos (**ANEXO 06**): valorar la idoneidad y pertinencia de la propuesta preliminar, verificando que no se haya omitido ninguna cuestión relevante, solicitar datos sobre la prioridad de las

⁶³ En la definición preliminar del modelo, la propuesta estaba formulada entendiendo que las dimensiones de la competencia evaluativa constituían en sí mismas competencias, por eso se empleó el plural.

competencias mostradas para reducir la propuesta inicial y resolver ciertas dudas sobre su posible estructuración y agrupación.

- **Análisis de las aportaciones.** El **ANEXO 07** recoge el informe resultante de esta quinta fase de análisis y detalla las principales conclusiones extraídas de la consulta. La dirección adoptada en las siguientes fases de trabajo ha estado determinada, en gran medida, por las orientaciones y sugerencias desprendidas de las mismas.

- **Redefinición del modelo de competencia evaluativa.** Este paso ha supuesto la redefinición de la propuesta preliminar según las directrices surgidas de la consulta a jueces. Ello ha conducido a la reordenación de la estructura general del modelo en los siguientes términos:

- ✓ *Reducción del número de competencias (dimensiones) inicialmente previstas.* De las diez formulaciones iniciales, la propuesta final acaba incorporando ocho. La reducción se produce por la supresión de la competencia décima, centrada en el modelado pedagógico, y por la fusión de las dos primeras (planificación estratégica de la evaluación e instrumentalización de la misma). A pesar de concentrar en sí claros desacuerdos, se opta por mantener la dimensión relativa a la atención a la diversidad en la evaluación para poder completar su exploración desde la perspectiva empírica.
- ✓ *Simplificación de las formulaciones* de las distintas dimensiones e integración de las sugerencias parciales a las mismas, en los términos expuestos en el informe (**ANEXO 07**).
- ✓ *Reducción de la complejidad de la propuesta en la medida de lo posible.* Esta recomendación se ha asumido como constante en el desarrollo de la nueva versión del modelo, aunque plantea nuevamente el difícil equilibrio entre elaborar una propuesta inabarcable y otra que simplifique en exceso la complejidad natural del proceso.

- **Creación de la rúbrica para su valoración.** Se trata de una de las recomendaciones derivadas de la fase anterior, aunque se representa como un paso más del proceso por la complejidad y la significatividad del cambio que ha supuesto su desarrollo. La tarea ha consistido en la articulación de una rúbrica evaluativa que precise y desarrolle indicadores y descriptores de logro para las ocho dimensiones resultantes en la propuesta revisada de la competencia evaluativa docente. Sobre la misma, realizamos a continuación algunas consideraciones:

- ✓ La definición de la rúbrica se ha realizado por medio de un proceso interno de revisión, reflexión y discusión, que ha dado lugar a cuatro versiones diferentes (ver figura 22). Su desarrollo se ha basado en las fuentes y referencias presentadas en el próximo apartado.
- ✓ La quinta versión de la rúbrica o rejilla es la que se ha adoptado como punto de partida para el diseño del marco empírico. Según se detalla en el capítulo séptimo, la misma da lugar a la configuración de dos cuestionarios para evaluar la competencia evaluativa docente con fines formativos: uno para el profesorado y otro para el colectivo de estudiantes. Para contribuir a su validación, por cada uno de los instrumentos se ha llevado a cabo una aplicación piloto. Las recomendaciones y sugerencias resultantes se han plasmado en una sexta versión de la rúbrica.
- ✓ Mostrada en el **ANEXO 05**, dicha versión es la que se asume como referencia en el transcurso de la investigación; en ella se concentra la propuesta de sistematización de la competencia evaluativa que se desarrolla en la segunda parte del capítulo. Aunque en este proceso de definición hubiera sido recomendable realizar una segunda consulta a jueces, la opción ha sido desestimada por cuestiones de viabilidad práctica. La dificultad de acceso a las/os expertas/os por un lado, y los tiempos de devolución de respuesta constatados en la primera consulta por el otro, así lo han condicionado.

El recorrido descrito da lugar a una propuesta tentativa, abierta a modificaciones y adecuaciones. De hecho, el análisis y discusión de los resultados asociados al estudio empírico, hace prever nuevas modificaciones sobre el modelo de competencia evaluativa diseñado y la rúbrica que lo operativiza.

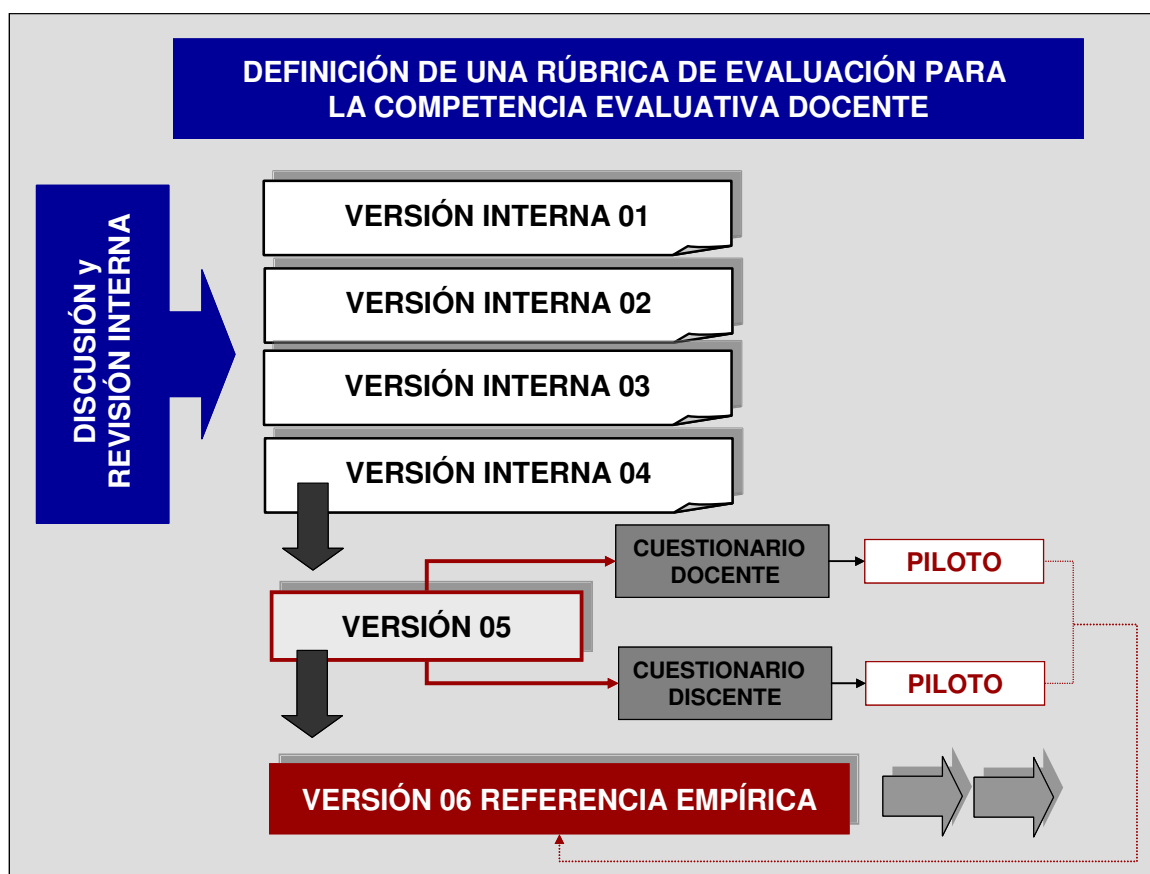


Figura 22. Definición de la rúbrica asociada al modelo de competencia evaluativa: descripción del proceso.

3. Fuentes consultadas en su definición

La propuesta ideada para pormenorizar la competencia evaluativa docente no es única ni absoluta, pero está fundamentada en la consulta a cuatro fuentes o referencias relevantes (ver figura 23). Desde ellas cobra sentido y se justifica la formulación de las ocho dimensiones presentadas y los indicadores y descriptores incorporados en su desarrollo:

- *Modelo ITC de evaluación del aprendizaje.* En coherencia con la propia estructura interna de la investigación, el primer referente para delimitar la competencia evaluativa del profesorado es la teoría evaluativa explicitada en el capítulo anterior. Basado, a su vez, en una selección amplia y significativa de referentes, el modelo ITC detalla una determinada manera de concebir la práctica evaluativa en la

universidad, particularmente en los grados vinculados a la formación básica de maestras/os.

- *Antecedentes o propuestas que han delimitado la competencia evaluativa docente.* Los trabajos que han tratado de sistematizar la competencia evaluativa del profesorado representan una segunda referencia o fuente de consulta.
- *Otras escalas o herramientas de probada validez que han atendido a la práctica evaluativa del profesorado en la formación universitaria.* La consulta de las escalas y herramientas empleadas en otros estudios sobre la evaluación del aprendizaje universitario representa otra de las referencias indagadas.
- *Especificaciones sobre la evaluación del aprendizaje incluidas en las escalas de valoración de la calidad docente administradas en los grados vinculados a la formación básica de maestras/os.* La consulta de las referencias anteriores se ha completado con un trabajo de campo directo sobre las escalas y cuestionarios empleados para valorar la calidad docente en las instituciones dedicadas a la formación básica de maestras/os de todo el estado.

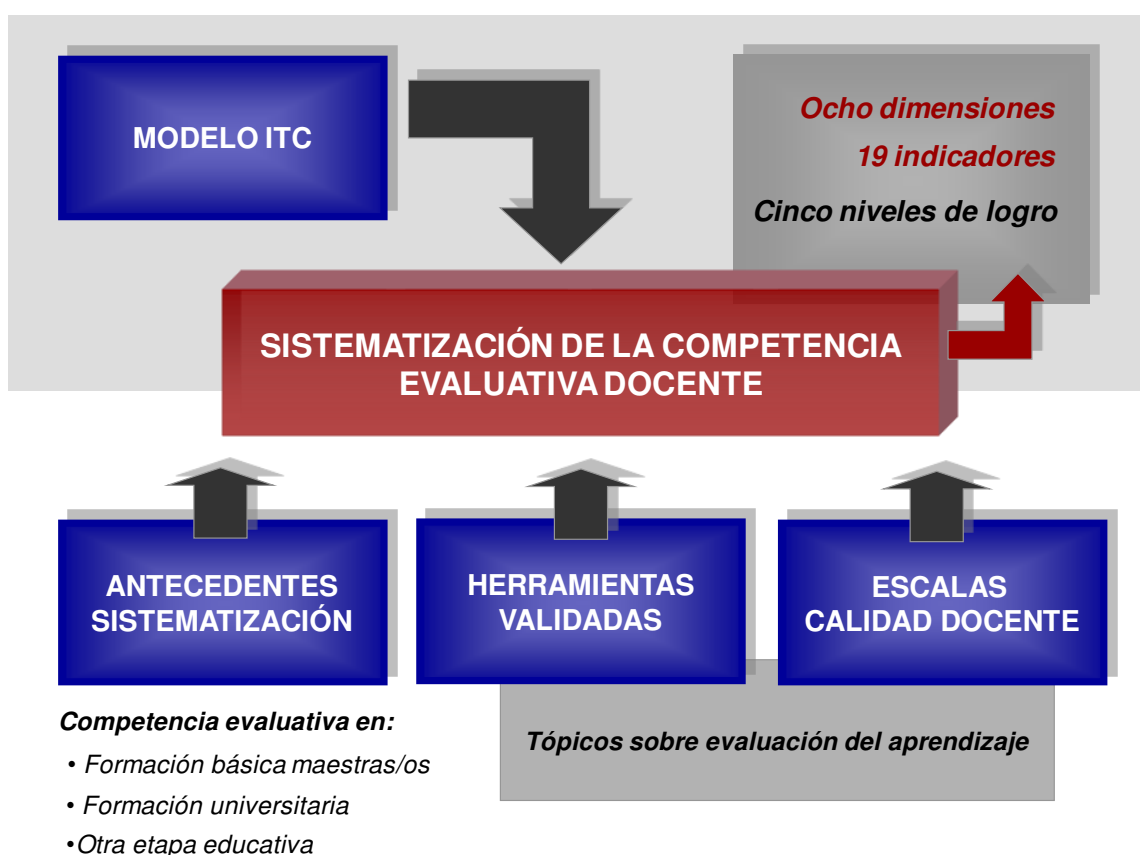


Figura 23. Referentes para la construcción de un modelo de sistematización de la competencia evaluativa docente.

A continuación se detallan las conclusiones extraídas de las tres últimas referencias, entendiendo que la presentación del modelo ITC ya ha sido atendida en el anterior capítulo.

3.1 Antecedentes: análisis de otras propuestas que han delimitado la competencia evaluativa docente.

Las propuestas que han tratado de delimitar los componentes de la competencia evaluativa del profesorado dedicado a la formación básica de maestras/os representan una referencia ineludible en la investigación en curso. Para favorecer la proximidad y la pertinencia de las fuentes consultadas, sólo se han seleccionado aquellas propuestas que comparten estrictamente nuestro objeto de estudio, dirigiendo su atención a la delimitación de la competencia evaluativa. En consecuencia, no se han incorporado a la consulta aquellas propuestas generales de competencias o estándares docentes que contemplan la evaluación del aprendizaje e incorporan alguna indicación para su definición y desarrollo⁶⁴.

La aplicación de dicho criterio reduce considerablemente el número de antecedentes a explorar. La búsqueda efectuada pone de manifiesto que, aunque la evaluación represente un objeto de investigación frecuente, son pocos los estudios que han abordado específicamente a la competencia requerida por el profesorado para su desempeño. Como recoge la tabla 45, en total se han considerado siete intentos previos de sistematización, con desigual grado de proximidad o coincidencia; entre ellos, uno se sitúa en la formación básica de maestras/os y otro más en la enseñanza universitaria. El resto de propuestas delimitan la competencia evaluativa para el profesorado de otras etapas educativas, principalmente en los niveles escolares.

⁶⁴ Entre las referencias descartadas se encuentran los estándares docentes del Arizona State Board for Education (ASBE, 1996) y los estándares docentes formulados por el Interstate New Teachers Assessment and Support Consortium (INTASC, 2011). En el contexto nacional, han quedado fuera de la selección de referentes las propuestas de Tejada (2009), Tejedor (2009), Mas (2012), Alegre y Villar (2004), V. Álvarez, García y Gil (1999) o Villa (2004). Centradas en la enseñanza en el perfil del profesorado universitario, todas ellas presentan la competencia evaluativa dentro del cuadro general de competencias docentes e incluyen alguna orientación específica para su desarrollo.

Tabla 45

Antecedentes en la sistematización de un modelo o teoría de la evaluación del aprendizaje

NIVEL	ANTECEDENTES: PROPUESTAS PREVIAS CONSIDERADAS
FORMACIÓN BÁSICA MAESTRAS/OS	• Toledo (2005, 2006)
FORMACIÓN UNIVERSITARIA	• Quesada et al. (2013)
ENSEÑANZA GENERAL	• American Federation of Teachers, National Council on Measurement in Education y National Education Association (1990) • American Association of School Administrators, National Association of Elementary School Principals, National Association of Secondary School Principals, National Council on Measurement in Education (1997) • Stiggins (1998) • J. Chappuis et al. (2012) • Monereo & Castelló (2009a)

Nota: La celda sombreada incluye un antecedente explorado en la fase de búsqueda confirmatoria; no se encontraba disponible en la fase inicial desarrollada para la delimitación de la propuesta teórica.

La consulta de referentes se ha realizado en dos momentos diferentes del proceso de investigación con distinta finalidad: antes de la formulación del modelo de competencia evaluativa (como fuente de consulta propiamente) y antes de dar por finalizada la redacción definitiva de la investigación (con fines confirmatorios). Sombreada en gris, la propuesta de Quesada et al. (2013) es la única incorporada a posteriori.

El modelo de competencia evaluativa elaborado integra las aportaciones procedentes de estas fuentes, ofreciendo una imagen amplia y completa de la misma. La tabla 46 recoge el reflejo de los siete antecedentes en las ocho dimensiones que componen la propuesta.

Tabla 46

Antecedentes en la sistematización de la competencia evaluativa del profesorado en la formación básica de maestras/os (FB), en la docencia universitaria (DU) o en la enseñanza general

	FB	DU	ENSEÑANZA GENERAL				
MODELO COMPETENCIA EVALUATIVA	1	2	3	4	5	6	7
1. DIMENSIÓN. Planificación estratégica de la evaluación del aprendizaje.	X	X	X	X	X	X	X
2. DIMENSIÓN. Clarificación de criterios y condiciones.	X				X		X
3. DIMENSIÓN. Atención a la diversidad en la evaluación del aprendizaje.	X						

MODELO COMPETENCIA EVALUATIVA	FB	DU	ENSEÑANZA GENERAL				
	1	2	3	4	5	6	7
4. DIMENSIÓN. <i>Implementación del sistema de evaluación.</i>			X	X			X
5. DIMENSIÓN. <i>Análisis e interpretación de las evidencias de aprendizaje.</i>	X		X	X	X		X
6. DIMENSIÓN. <i>Regulación del aprendizaje y de la enseñanza.</i>	X	X	X	X	X	X	X
7. DIMENSIÓN. <i>Participación discente en la evaluación del aprendizaje.</i>	X	X			X	X	
8. DIMENSIÓN. <i>Revisión del sistema de evaluación.</i>		X	X	X			X

Nota: Las siete columnas de la derecha recogen los antecedentes consultados en el mismo orden de presentación: (1) Toledo (2005, 2006); (2) Quesada et al. (2013); (3) American Federation of Teachers, National Council on Measurement in Education, y National Education Association (1990); (4) American Association of School Administrators, National Association of Elementary School Principals, National Association of Secondary School Principals, National Council on Measurement in Education (1997); (5) Stiggins (1998); (6) J. Chappuis et al. (2012); (7) Monereo & Castelló (2009a).

La segunda columna aparece sombreada en gris por haber sido incorporada en la fase final de confirmación de fuentes.

Los próximos apartados describen brevemente cada uno de los tres bloques de antecedentes. Su consideración permite comprender la referencia a los mismos en la justificación de las dimensiones de la competencia evaluativa presentada al final del capítulo.

3.1.1 Competencia evaluativa: delimitación para el profesorado universitario de los grados vinculados a la formación básica de maestras/os de Educación Infantil y/o Primaria.

La conclusión más destacable de este primer nivel de consulta es que no hemos hallado ninguna propuesta concreta para precisar la competencia evaluativa docente en el espacio formativo acotado. Varios estudios que se han aproximado a la evaluación del aprendizaje en la formación básica de maestras/os o profesorado (entre otros, Arribas et al., 2010; Hamodi & López, 2012; Hortigüela et al., 2015; V. M. López & Palacios, 2012; Palacios & López, 2013; Ruiz-Gallardo et al. 2013; Smitter, 2006; Smitter & Siso, 2007), pero ninguno de ellos ha entrado a delimitar la competencia evaluativa requerida por el profesorado.

La referencia más próxima la ofrece Toledo (2005), quien formula una batería *de 23 indicadores para el uso formativo de la evaluación del aprendizaje entre el profesorado universitario dedicado a la formación básica de maestras/os*. En un trabajo posterior realiza

una *descripción de la competencia evaluativa del profesorado* a partir de los mismos (Toledo, 2006).

3.1.2 Competencia evaluativa: delimitación para el profesorado universitario.

Ampliando el enfoque de la búsqueda, en una segunda aproximación se han considerado las propuestas centradas en la delimitación de la competencia evaluativa del profesorado dedicado a la enseñanza universitaria. A este nivel vuelve a constatar que la sistematización de la actuación evaluativa no representa un foco de atención frecuente. Esta circunstancia dificulta la fase inicial de consulta, pero avala la pertinencia y necesidad de modelos específicos que orienten procesos de formación o revisión en materia evaluativa.

En la primera fase de consulta, la búsqueda de propuestas centradas en la delimitación de la competencia evaluativa del profesorado universitario no produjo resultados. Bien avanzado el marco empírico de la investigación, en la fase confirmatoria se produjo la publicación del trabajo de Quesada et al. (2013) sobre la *delimitación de funciones evaluativas del profesorado universitario*. Esta contribución tiene lugar dentro del proceso de construcción de ActEval, una herramienta para la evaluación de la actividad evaluadora referida en el siguiente bloque de referentes. Parte de la enumeración de cuatro grandes funciones vinculadas al ejercicio de la evaluación del aprendizaje discente en la docencia universitaria. Como refleja la síntesis presentada en la tabla 46, todas ellas están contempladas en las dimensiones del modelo propuesto, reforzando su adecuación y pertinencia.

3.1.3 Competencia evaluativa: delimitación para el profesorado.

El tercer bloque de antecedentes lo conforman las propuestas que han sistematizado la competencia evaluativa del profesorado fuera del contexto formativo universitario o sin referencia expresa al mismo. Entre ellas se cuentan cinco referencias, todas ellas seleccionadas en la fase inicial de búsqueda:

- *Estándares para la competencia docente en la evaluación educativa de las/os estudiantes* (American Federation of Teachers, National Council on Measurement in Education, & National Education Association, 1990). Estos estándares recogen el

producto de un proceso de trabajo conjunto iniciado en 1987 por las asociaciones profesionales firmantes. Responde a la inquietud por las deficiencias acusadas en el uso de la evaluación del aprendizaje discente y su impacto en la efectividad del sistema educativo. La propuesta identifica siete estándares asociados a la competencia evaluativa de las/os maestras/os de niveles escolares e incorpora una breve descripción de los mismos. Su objetivo es orientar la práctica evaluativa del profesorado y ofrecer un marco de referencia a las/os responsables del diseño de acciones de formación profesional inicial o permanente.

A pesar de constituir una referencia lejana en el tiempo e indirecta por su ámbito de aplicación, recopila orientaciones válidas y pertinentes para definir la competencia evaluativa docente. En suma, plantea una propuesta teórica y descriptiva, que integra buena parte de los elementos clave identificados para el ejercicio de la evaluación del aprendizaje en el modelo ITC (ver tabla 46); en consecuencia, guarda una correspondencia razonable con las dimensiones de la competencia evaluativa delimitadas a partir del mismo.

- *Estándares de competencia en evaluación de estudiantes para administradores de la educación* (American Association of School Administrators, National Association of Elementary School Principals, National Association of Secondary School Principals, & National Council on Measurement in Education, 1997). Esta segunda propuesta está ideada para los agentes administradores de la educación, más que para el profesorado en ejercicio. Su objetivo es guiar procesos de formación inicial o permanente de los administradores de la educación, de modo que éstos puedan ejercer con solvencia las tres principales funciones que se les encomiendan con respecto a la evaluación: ayudar a las/os maestras/os a usar la evaluación de manera efectiva, liderar la creación y el desarrollo de políticas evaluativas y emplear los resultados constatados en la toma de decisiones sobre alumnado, profesorado o instrucción.

La propuesta incorpora un total de doce estándares, clasificados según la función específica a la que sirven. Los dos primeros, referidos a la promoción de las prácticas evaluativas efectivas entre el profesorado, remiten directamente a los *Estándares para la competencia docente en la evaluación educativa de las/os estudiantes* incluidos en la referencia anterior⁶⁵. Los diez restantes, hacen referencia a tareas o responsabilidades que superan el ámbito de incumbencia

⁶⁵ Según se expone en el propio documento de presentación, las/os administradores deben dominar las competencias evaluativas orientadas al profesorado para comprenderlas en profundidad y poder supervisarlas.

docente, alcanzando las responsabilidades de los agentes administradores con respecto a la política evaluativa y a la toma de decisiones.

- *Auditing the quality of classroom assessment training in teacher education programs* (Stiggins, 1998). Esta referencia identifica una serie de criterios para auditar la calidad de la preparación o capacitación evaluativa que se propicia en los centros de formación básica del profesorado. Aunque la propuesta se genera en el mismo contexto formativo en el que se sitúa la presente investigación, su contribución hace referencia a la competencia evaluativa de las/os futuras/os maestras/os, no a la del profesorado universitario. Como resume el propio Stiggins (1998), la motivación que inspira su trabajo es determinar si se aportan o no al mercado laboral y al sistema educativo maestras/os adecuadamente formadas/os para afrontar los retos de la evaluación.

Los criterios incluidos en la propuesta surgen de la revisión y ajuste de los *Estándares para la competencia docente en la evaluación educativa de las/os estudiantes* (American Association of School Administrators, National Association of Elementary School Principals, National Association of Secondary School Principals, & National Council on Measurement in Education, 1997). Desde ese marco de referencia, incorpora siete líneas que deberían atenderse en la formación evaluativa básica, por constituir los grandes ejes de mejora para la misma. Su propuesta surge del trabajo de identificación de tareas evaluativas realizado por el *Assessment Training Institute* con cientos de distritos escolares y miles de profesoras/es en ejercicio.

- *Indicadores de una buena práctica evaluativa en el aula* (J. Chappuis et al., 2012). Esta cuarta referencia está ligada a las cinco claves y las siete estrategias de la evaluación de calidad consideradas en el capítulo cuarto para la delimitación del modelo ITC (Stiggins et al., 2006). Las/os autores completan su propuesta formulando indicadores de buena praxis para el profesorado en cada una de las cinco claves. En total, conforman un sistema de 23 indicadores que ofrece una descripción indirecta de la competencia evaluativa docente.

Ubicada en el contexto escolar del Reino Unido, esta referencia se adscribe a la conceptualización teórica de la evaluación para el aprendizaje, lo que explica su especial insistencia en las dimensiones referidas a la planificación evaluativa, a la regulación del aprendizaje y a la participación discente (tabla 46).

- *Competencia docente en evaluación del aprendizaje* (Monereo & Castelló, 2009a). Este último antecedente surge de la revisión e integración de dos propuestas que ya han sido referidas en el presente trabajo: la de la American Federation of Teachers, el National Council on Measurement in Education y la National Education Association (1990) y la enunciada por el ASBE⁶⁶ (Arizona's State Board for Education, 1996). El resultado es una propuesta conformada por cinco macro competencias evaluativas clave y una breve descripción de lo que supone el ejercicio de cada una de ellas. Muy próxima a las dimensiones delimitadas en el modelo de competencia evaluativa presentado en el cuarto apartado, ofrece una orientación general sin adscribirse a ninguna etapa educativa en particular.

3.2 Otras escalas o herramientas de probada validez que han atendido a la práctica evaluativa del profesorado en la formación universitaria.

Las herramientas empleadas con probada validez en estudios sobre la práctica evaluativa en la universidad representan otra de las referencias consideradas en la delimitación de la competencia evaluativa docente (figura 23).

En total, se han analizado 30 instrumentos, utilizados en investigaciones nacionales e internacionales a lo largo de las dos últimas décadas. Aunque ninguno de ellos esté expresamente ideado para valorar la competencia evaluativa del profesorado, cuestionan cómo se produce la evaluación del aprendizaje docente en la universidad desde distintas perspectivas; en consecuencia, proporcionan una medida indirecta de los aspectos a los que se confiere relevancia en el proceso evaluativo, y en los que el profesorado al cargo debería mostrar cierta solvencia.

La tabla 47 detalla cuáles son las escalas y cuestionarios considerados y los clasifica de acuerdo a dos criterios:

- *El tipo de estudio en el que han sido empleados.* Los cuestionarios y escalas recopiladas están incluidos en dos tipos de estudio a los que denominaremos específicos y generales, según su vinculación con la temática evaluativa. En el

⁶⁶ Como ya ha sido justificado, esta referencia ha sido excluida de los referentes, porque se trata de un modelo de estándares docentes, no de una propuesta específicamente dirigida a profundizar en la competencia evaluativa. Incluye un estándar completo dedicado a la evaluación y varias alusiones transversales al proceso.

primer grupo figuran los instrumentos aplicados en estudios específicamente empleados para indagar sobre la actividad evaluativa en la universidad. Frente a ellos, el segundo grupo lo constituyen las escalas y cuestionarios orientados a la valoración de la actividad docente, con al menos una dimensión referida a la evaluación del aprendizaje del alumnado.

- *La fuente a la que se dirigen.* Generales o específicos, los instrumentos recopilados en ambos grupos están orientados a captar perspectivas diferentes sobre la evaluación del aprendizaje, según los propósitos del estudio al que sirvan. Dependiendo de la fuente de la que pretendan recabar información la clasificación realizada discrimina entre las escalas o cuestionarios para estudiantes, para profesorado o aquellos otros ideados en versiones diferentes para captar ambas perspectivas.

Como en el análisis de los antecedentes, la búsqueda de referencias en cuestionarios y herramientas se ha llevado a cabo antes y después del estudio empírico, como fuente de consulta y como recurso de confirmación y actualización, respectivamente. En suma, los instrumentos revisados se aproximan al fenómeno evaluativo desde perspectivas diferentes, con distintos propósitos y en desigual grado de profundidad. En consecuencia, sobre los mismos se ha efectuado un análisis general, atendiendo a dos cuestiones principalmente: el contenido evaluativo sondeado y el formato de indagación empleado.

Tabla 47

Referentes para la configuración del modelo de competencia evaluativa docente: escalas y cuestionarios aplicados al análisis de las prácticas evaluativas en la universidad

TIPO DE ESTUDIO	PERSPECTIVA	HERRAMIENTA
ESPECÍFICO	ESTUDIANTES	• <i>M. Fernández (1989).</i> • <i>Flórez (1999).</i> • <i>Assessment Experience Questionnaire (AEQ) (Gibbs & Simpson, 2003)⁶⁷;</i>
		• <i>Trillo & Porto (1999)</i> • <i>Porto & Rodríguez (2009)⁶⁸</i> • <i>A. Gargallo (2009)</i>
• <i>Gil-Flores (2012)</i> • <i>V. M. López & Palacios (2012)</i>		
ACTIVIDAD EVALUADORA		

⁶⁷ Recientemente dicha herramienta ha sido validada por J. L. Núñez y Reyes (2014) para su uso en España.

⁶⁸ Plantea modificaciones sobre el cuestionario empleado en Trillo y Porto (1999).

TIPO DE ESTUDIO	PERSPECTIVA	HERRAMIENTA
ESPECÍFICO	DOCENTES	• <i>ActEval</i> (Quesada et al., 2013). • <i>Palacios & López</i> (2013)
	DOCENTES + ESTUDIANTES	• <i>Tejedor</i> (1999) • <i>MacLellan</i> (2001) • <i>L. F. Martínez et al.</i> (2012) • <i>G. Rodríguez et al.</i> (2012) • <i>Ruiz-Gallardo et al.</i> (2013).
GENERAL	DOCENTES + ESTUDIANTES	• <i>Tejedor</i> (2003) • <i>Arribas et al.</i> (2010)⁶⁹ • <i>Gutiérrez et al.</i> (2011)
	DOCENTES	• <i>L. Prieto</i> (2007) • <i>V. Álvarez et al.</i> (2011) • <i>Cuestionario de Evaluación de la Metodología Docente y Evaluativa de los Profesores Universitarios (CEMEDEPU)</i> (B. Gargallo et al., 2011) • <i>B. García et al.</i> (2012)
	ESTUDIANTES	• <i>Course Experience Questionnaire (CEQ)</i> (Ramsden, 1991) • <i>Cuestionario de Evaluación Docente por el Alumno (CEDA)</i> (J. M. García, 1997) • <i>V. Álvarez et al.</i> (1999) • <i>Rizo</i>⁷⁰ (1999) • <i>Rué</i> (2009b) • <i>B. Gargallo et al.</i> (2010) • <i>National Student Survey (NSS)</i> (Buckley, 2012) • <i>Castejón, Santos, & Palacios</i> (en prensa)

Nota: Las celdas sombreadas contienen instrumentos explorados en la fase de búsqueda confirmatoria, publicadas a partir de 2011; no se encontraban disponibles en la fase inicial desarrollada para la delimitación de la propuesta teórica.

En cuanto al contenido, los instrumentos consultados ofrecen una visión de la evaluación del aprendizaje muy centrada en su vertiente instrumental. La recogida de información se centra principalmente en el tipo de procedimientos empleados (exámenes de diverso tipo, trabajos, etc.), en la frecuencia y los momentos en los que tiene lugar la evaluación, y en la transparencia de los criterios y procesos desarrollados. También incluyen alusiones puntuales a la fiabilidad de la evaluación, entendida como precisión y objetividad en sus conclusiones. Todos estos aspectos se reflejan principalmente en cuatro de las dimensiones previstas en el modelo de competencia evaluativa: en la planificación

⁶⁹ También estudiantes egresados.

⁷⁰ Se trata de un cuestionario de competencia docente orientado a varias audiencias. La parte de la evaluación la formula al alumnado, de ahí que se sitúe en la perspectiva discente.

estratégica de la evaluación, en la clarificación de criterios y condiciones, en la implementación coherente y en la dedicada al análisis e interpretación de evidencias (la quinta).

En el extremo opuesto, los instrumentos revisados no incorporan referencias a la revisión de la evaluación ni al tratamiento de la diversidad en su desarrollo (dimensiones octava y tercera del modelo). Aspectos como la participación del alumnado o las contribuciones formativas de la evaluación al aprendizaje sólo aparecen de forma puntual en la muestra de herramientas.

Frente a esta visión instrumental que devuelven los recursos revisados, la presente investigación ofrece una perspectiva más amplia de la evaluación; más que recabar información concreta sobre su continuidad o el tipo de instrumentos o procedimientos empleados, atiende al planteamiento general que se hace de la misma en el sistema didáctico de un módulo/asignatura determinada.

En cuanto al formato de indagación, impera el empleo de recursos económicos basados en cuestionarios o escalas tipo Likert. Orientados a estudiantes y/o profesorado, coinciden en plantear una serie de tópicos evaluativos y un repertorio cerrado de alternativas de respuesta, por lo general, vinculadas al grado de acuerdo o frecuencia de cumplimiento. En claro contraste, la investigación en curso recurre al formato de rúbrica o rejilla de evaluación para operativizar el modelo de competencia evaluativa elaborado.

3.3 Especificaciones sobre la competencia evaluativa incluidas en las escalas de evaluación de la calidad docente administradas en los grados vinculados a la formación básica de maestras/os.

Esta última referencia para la delimitación del modelo de competencia evaluativa docente (figura 23) es el resultado de un trabajo de campo desarrollado en el transcurso de la investigación principal. En él se ha efectuado un análisis de las escalas y cuestionarios aplicados para la evaluación de la calidad docente en las titulaciones asociadas a la formación básica de maestras/os en todo el territorio español.

Esta indagación se produce por la convicción de que tras la evaluación de la docencia que se efectúa en cada caso, reposa latente el modelo de calidad teórico al que se desea tender (Ramsden, 1991; Rizo, 1999; Tejedor, 2003). Desde esta perspectiva, el

estudio se proyecta para identificar los indicadores evaluativos presentes en dichas herramientas, asumido que se consideran elementos relevantes y de reconocido impacto en la calidad general de la enseñanza.

Los siguientes apartados resumen el planteamiento metodológico del trabajo de campo efectuado y sus principales conclusiones.

3.3.1 Descripción del trabajo de campo.

El **objetivo** del estudio desarrollado ha sido analizar la presencia de tópicos referidos a la evaluación del aprendizaje en las escalas de valoración de la calidad docente administradas a las/os estudiantes en centros universitarios vinculados a la formación básica de maestra/os. El trabajo de campo se ha efectuado para respaldar la formulación teórica del modelo de competencia evaluativa sobre el que se desarrolla el cuerpo empírico de la investigación.

La **población** del estudio la han constituido las escalas de evaluación de la calidad docente administradas a las/os estudiantes en centros universitarios que ofrecen los Grados de Educación Infantil y Educación Primaria (o equivalentes) en todo el estado. Se ha optado por acotar la recogida a los instrumentos destinados al alumnado por ser los más frecuentes y extendidos en las universidades españolas (Berk, 2005; De Juan, Pérez, Gómez-Torre, Vizcaya, & Mora, 2007; Murillo, 2008; S. Rodríguez, 2003). De acuerdo con las directrices que se desprenden del programa DOCENTIA para la evaluación de la calidad docente a nivel nacional (ANECA, 2006), las/os estudiantes representan una fuente destacada de consulta y el cuestionario la vía preferente de acceso a la misma. Para la consulta del punto de vista del profesorado y las/os responsables académicos se recomienda el empleo de otro tipo de procedimientos de informe o auto-informe, menos accesibles a la investigación externa.

En España hay 63 universidades implicadas en la formación básica de maestras/os de Educación Infantil y Primaria, siendo el 65,07% de las mismas públicas y de titularidad privada el 34,92% restante. Estas universidades imparten ambas titulaciones en un total de 97 centros de enseñanza superior repartidos por todo el estado: el 56,70% son centros públicos, el 24,74% de titularidad privada y adscritos los restantes. La distribución de estos centros y universidades en las 19 comunidades autónomas se representa de forma sintética en el **ANEXO 08**.

La **muestra** del estudio está conformada por un total de 63 escalas o cuestionarios para la evaluación de la calidad docente en la versión destinada al colectivo discente (**ANEXO 09**). Los instrumentos han sido recopilados por tres vías complementarias:

- *Escalas aportadas en un estudio anterior.* El 9,52% de las escalas analizadas proceden de un proyecto de investigación coordinado por Marbán (2012), en el que nueve universidades españolas se implican en el análisis de escalas de evaluación de la calidad docente desde la perspectiva de los estándares internacionales de excelencia.
- *Procedimiento de solicitud directa dirigido a los centros universitarios incluidos en la población.* La solicitud se ha efectuado por correo electrónico a las/os directoras/es de centros y/o decanas/os de facultad. Esta vía ha dado lugar a la recopilación de un total de 23 escalas (36,50% de todas las que conforman la muestra inicial).
- *Análisis del contenido publicado en las páginas web.* Una tercera forma de acceder a las escalas de valoración de la calidad docente ha sido recurrir directamente a la información institucional presentada vía web. Unido a la implementación del programa DOCENTIA de la ANECA (2006) y a la proliferación de los vicerrectorados, secretarías y unidades técnicas de calidad, son muchas las universidades y centros que reservan un espacio de su página web para publicar su sistema de gestión de garantía interna de la calidad y las herramientas de recogida de evidencias que incorporan. Por esta vía han sido recopiladas 34 escalas (el 53,96% de la muestra inicial).

En la muestra conformada quedan representadas el 93% de las universidades que ofrecen los grados acotados en el estado, el 98% de las públicas y el 74% de las de titularidad privada. La representación por centro es del 77% (81% de los centros públicos, 64% de los privados, 17 69% de los centros privados adscritos a universidades públicas y el 100% de los adscritos a universidades privadas)⁷¹ (**ANEXO 10**).

Los primeros análisis sobre la configuración de la muestra conducen a las siguientes consideraciones:

⁷¹ Estos datos han sido calculados considerando sólo el número de escalas directamente recopiladas de los diversos tipos de centro. Los porcentajes de representación real son mayores que los reflejados; como se detalla en la columna de observaciones del **ANEXO 09** (apelaciones a la “Representación indirecta”), en ocasiones se produce confirmación de que una misma escala recogida de una universidad o centro (un caso en la muestra), se aplica idénticamente en otros centros que no participan en la investigación (representación indirecta).

- Los centros dependientes de una misma institución universitaria, por lo general aplican la misma herramienta para la evaluación de la calidad docente. Para evitar la distorsión de los resultados, en casos de repetición constatada, el análisis ha sido efectuado contando sólo con una escala por universidad. Esta medida ha supuesto considerar en el análisis 57 escalas, el 90,47% de las recopiladas inicialmente. Los centros dependientes de una misma universidad que administran herramientas diferentes han sido considerados como casos independientes de la muestra. Esta información se detalla en la relación total de centros participantes del **ANEXO 09**.
- En la muestra final se integran tres escalas con formato diferente pero idéntico contenido⁷². Son las correspondientes a las Universidades de Huelva, Almería y Cádiz. En la misma comunidad autónoma, las escalas aplicadas en las Universidades de Granada y Córdoba también muestran cierta similitud. A pesar de la coincidencia, estos casos se han mantenido en la muestra por representar a universidades diferentes.
- Todas las escalas recopiladas son de carácter general. Es decir, se aplican en una misma universidad o centro para la evaluación de la calidad docente en las distintas titulaciones impartidas. No se ha dado cuenta de ninguna específicamente diseñada o adecuada para su aplicación en las titulaciones asociadas a la formación básica de maestras/os.

3.3.2 Principales conclusiones⁷³.

El análisis de los instrumentos de evaluación recopilados ha dado lugar a la identificación y transcripción inicial de 150 cuestiones sobre la evaluación, procedentes de 53 de las 57 escalas integradas en la muestra descrita. Con un nivel de dispersión por encima de los 8 puntos ($s = 8,33$), el promedio de ítems que componen las herramientas analizadas es de 19,38. La concentración media de tópicos evaluativos por escala es de 2,63 ítems, aunque se acusan importantes diferencias dentro de la muestra ($s = 1,519$; *Moda* = 2).

⁷² La coincidencia se debe presumiblemente a las orientaciones de la Agencia Andaluza del Conocimiento (AAC), que remiten a la referencia común que proporciona la ANECA (2006) para el diseño de los modelos de evaluación institucionales.

⁷³ Está pendiente una publicación más profunda y extensa sobre el estudio realizado y sus principales conclusiones, de modo que sólo se comunican las aportaciones más relevantes para la configuración de un modelo que sistematice la competencia evaluativa docente.

El análisis cualitativo de los 150 ítems o tópicos que refieren a la evaluación del aprendizaje ha dado lugar a la identificación de 15 categorías o núcleos temáticos estrechamente interrelacionados entre sí (ver figura 24). Cada una de las categorías recoge una cuestión específica en relación con la evaluación del aprendizaje, de modo que el sistema que a continuación se describe agrupa por afinidad los ítems identificados (**ANEXO 11**):

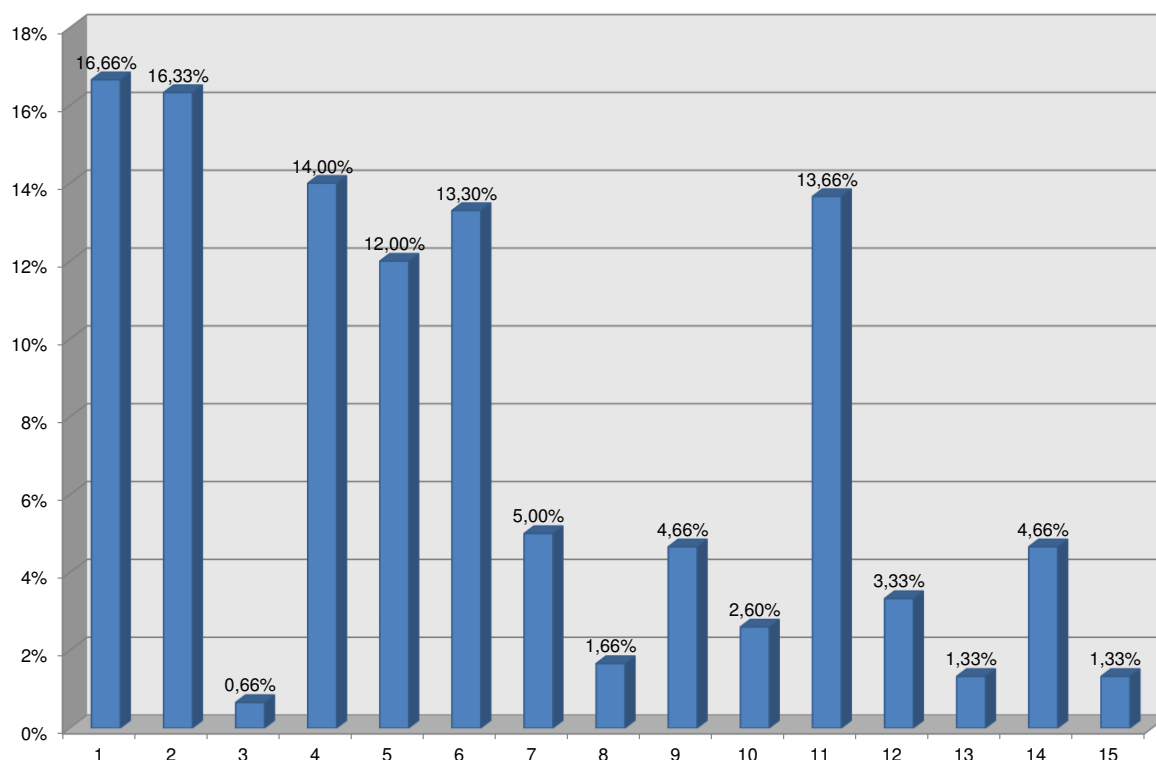


Figura 24. Representación de las 15 categorías temáticas identificadas en el banco de tópicos evaluativos.

1	<i>Información general</i>	9	<i>Variedad de objetos e instrumentos</i>
2	<i>Información específica</i>	10	<i>Transparencia del proceso</i>
3	<i>Planificación</i>	11	<i>Adecuación y validez</i>
4	<i>Desarrollo coherente</i>	12	<i>Justicia y precisión</i>
5	<i>Alineación</i>	13	<i>Participación del alumnado</i>
6	<i>Evaluación de competencias</i>	14	<i>Nivel de exigencia</i>
7	<i>Orientación formativa</i>	15	<i>Satisfacción con la evaluación</i>
8	<i>Evaluación continua</i>		

- **Información general previa.** La categoría que más se repite en las escalas consultadas es la que asocia la evaluación a la información general que se ofrece sobre el módulo/ asignatura a las/os estudiantes al comienzo del proceso formativo (a priori). Estrechamente vinculados al programa o la guía docente, son enunciados en los que la evaluación del aprendizaje aparece junto a otros elementos clave que

caracterizan el trabajo a efectuar (por lo general, competencias y objetivos, contenidos, metodología y, en algunos casos, bibliografía de referencia).

- **Información específica previa.** Frente a la información general referida en la categoría anterior, se evidencia un segundo grupo de ítems que apela directamente a la información proporcionada al alumnado, al comienzo de su proceso formativo, sobre la evaluación del aprendizaje. Como se observa en el **ANEXO 11**, incluye enunciados en los que se hace alusión expresa a la información que se ofrece a las/os estudiantes sobre los criterios y procedimientos de evaluación que serán aplicados en un módulo o asignatura determinada.
- **Planificación.** La tercera categoría se constituye de un único ítem recogido de forma aislada en una escala de las 53 que incluyen aportaciones en este sentido. Como se desprende de su propio enunciado, hace referencia a la adecuada planificación de las actividades evaluativas, en paralelo a la actividad docente.
- **Desarrollo coherente.** Entre los 150 tópicos evaluativos recabados de la muestra se distingue un cuarto núcleo denominado desarrollo coherente. Los ítems que se concentran en el mismo cuestionan el grado de cumplimiento práctico que percibe el alumnado de los compromisos evaluativos informados a priori.
- **Alineación.** El análisis de los tópicos evaluativos seleccionados conduce a la identificación de una quinta categoría denominada alineación en terminología de Biggs (2005). Quedan recogidos en la misma, ítems que plantean la relación de coherencia e interdependencia de la evaluación con el resto de elementos curriculares, en especial, con los objetivos de aprendizaje establecidos y con las actividades desempeñadas para su logro.
- **Evaluación de competencias.** La evaluación de competencias representa la sexta categoría evidenciada entre los tópicos que hacen referencia a la evaluación del aprendizaje. Incluye ítems referidos al objeto de la evaluación, en los que se preconiza la atención a competencias, frente a la clásica perspectiva centrada en la valoración de contenidos.
- **Orientación formativa o pedagógica.** La orientación pedagógica de la evaluación representa una séptima categoría temática dentro del banco de ítems evaluativos. Los tópicos que en él se incluyen cuestionan la contribución de la evaluación al proceso de aprendizaje, presentándola como un recurso que informa del avance del alumnado hacia los logros previstos. El tiempo verbal y la terminología empleada en su redacción los aleja de la evaluación final y los aproxima a la próxima categoría (**ANEXO 11**).

- **Evaluación continua.** Frente a la clásica evaluación final, los ítems incluidos en este núcleo hacen referencia expresa a la evaluación continua, aquella que no se desarrolla de forma puntual y aislada.
- **Variedad de objetos e instrumentos de evaluación.** Este núcleo reúne ítems que plantean la variedad de técnicas y objetos de evaluación. Expuestos en el **ANEXO 11**, los tópicos incluidos en él plantean el uso de instrumentos de evaluación diversos (exámenes, exposiciones, trabajos de distinto tipo, portafolios, etc.) para la consideración de distintos aspectos (no sólo contenidos conceptuales y elementos examinables: aportación y participación en clase, progreso, asistencia, etc.).
- **Transparencia del proceso evaluativo.** La transparencia con la que transcurre el proceso de evaluación representa otro núcleo temático reflejado en las escalas consultadas. En ella destaca la corrección de la evaluación como rasgo deseable, asociada a la publicidad y explicitación de los criterios en los que se apoya y a su aplicación fiel.
- **Adecuación y validez.** En esta nueva categoría se agrupan formulaciones que ponen el énfasis en que la evaluación se lleve a cabo por medio de procedimientos e instrumentos adecuados, capaces de captar sin desviaciones el aprendizaje que pretende valorarse.
- **Justicia y precisión.** La precisión y la justicia de los procedimientos de evaluación del aprendizaje, representa otra cuestión valorada en las escalas de evaluación docente analizadas. En esta categoría, se registran tópicos en los que se plantea al alumnado la justicia de la evaluación, en términos de rigor o precisión (fiabilidad).
- **Participación del alumnado en la evaluación.** Aunque de forma minoritaria, la participación de las/os estudiantes en los procesos de evaluación del aprendizaje también figura en la muestra de ítems específicos sobre el tema. Esta participación se contempla en dos niveles: en el desarrollo de prácticas de autoevaluación y/o coevaluación y en la fase de diseño de los propios procedimientos de evaluación.
- **Nivel de exigencia.** El nivel de exigencia que perciben las/os estudiantes en la evaluación, representa otra de las categorías identificadas tras el análisis cualitativo de los tópicos. Estrechamente ligados a los procesos de calificación, este núcleo incorpora ítems que cuestionan el ajuste entre el nivel de dificultad o exigencia atribuido a la evaluación y los requerimientos del módulo/ asignatura. Para expresarlos, apelan en su mayoría a la carga de créditos ECTS o directamente a los aprendizajes pretendidos.

- **Satisfacción del alumnado con la evaluación.** La última categoría discriminada entre los tópicos evaluativos hace referencia a la satisfacción general que manifiesta el alumnado con el sistema de evaluación experimentado.

Siguiendo a B. Álvarez et al. (2007) la evaluación del aprendizaje constituye una de las dimensiones más representada en los procesos de evaluación de la calidad docente y, más concretamente, en los instrumentos que éstos ponen en juego. El estudio efectuado sólo se ha centrado en los ítems evaluativos, de modo que no puede alcanzar conclusiones comparativas en este sentido. No obstante, el promedio de representación calculado para la evaluación del aprendizaje en estas escalas (el 13,57%⁷⁴) no parece difícilmente superable por otros aspectos asociados a la docencia (metodología y recursos, relación con el alumnado o dominio de la materia y entusiasmo docente, entre otras). Los resultados puestos de manifiesto por I. González y López (2010) en un análisis similar refuerzan tal suposición.

Al margen de su representación global, los 15 temas reflejados en los instrumentos ofrecen una imagen indirecta de aquellas cuestiones relacionadas con la evaluación a las que institucionalmente se concede mayor atención y relevancia. Desde esta perspectiva, el modelo de competencia evaluativa propuesto incorpora en su desarrollo los rasgos y componentes destacados en este análisis (ver tabla 48):

Tabla 48
Representación de los principales tópicos evaluativos en las dimensiones integrantes del modelo

TÓPICOS EVALUATIVOS	MODELO DE COMPETENCIA EVALUATIVA: DIMENSIONES							
	1	2	3	4	5	6	7	8
INFORMACIÓN GENERAL		X						
INFORMACIÓN ESPECÍFICA		X						
PLANIFICACIÓN	X							
DESARROLLO COHERENTE				X				
ALINEACIÓN	X							
EVALUACIÓN DE COMPETENCIAS	X							
ORIENTACIÓN FORMATIVA	X					X		
EVALUACIÓN CONTINUA	X							
VARIEDAD DE OBJETOS E INSTRUMENTOS	X							
TRANSPARENCIA DEL PROCESO		X						X

⁷⁴ Estimación calculada sobre el promedio total de ítems registrados en la muestra ($M = 19,38$) y la concentración media de tópicos evaluativos dentro de las mismas ($M = 2,63$).

TÓPICOS EVALUATIVOS	MODELO DE COMPETENCIA EVALUATIVA: DIMENSIONES							
	1	2	3	4	5	6	7	8
ADECUACIÓN Y VALIDEZ					X			X
JUSTICIA Y PRECISIÓN					X			X
PARTICIPACIÓN DEL ALUMNADO							X	
NIVEL DE EXIGENCIA	X							
SATISFACCIÓN CON LA EVALUACIÓN								X

Nota: Las dimensiones son: (1) Planificación estratégica de la evaluación del aprendizaje; (2) Clarificación de criterios y condiciones; (3) Atención a la diversidad a la evaluación del aprendizaje; (4) Implementación del sistema de evaluación; (5) Análisis e interpretación de las evidencias de aprendizaje; (6) Regulación del aprendizaje y de la enseñanza; (7) Participación discente en la evaluación del aprendizaje; (8) Revisión del sistema de evaluación.

Las conclusiones extraídas de la indagación pueden ser confrontadas con cuatro referentes de interés que ofrecen una aproximación complementaria:

- El Programa DOCENTIA (ANECA, 2006) ofrece a nivel estatal un modelo de evaluación para orientar a las instituciones universitarias en el diseño y desarrollo de programas de evaluación docente. Para ello, propone el análisis de tres grandes dimensiones vinculadas a la enseñanza universitaria y una serie de aspectos concretos dentro de cada una de ellas. Entre estas referencias, se cuentan tres menciones expresas a la evaluación del aprendizaje, todas ellas contempladas en la categorización propuesta: la planificación de criterios y métodos de evaluación, el desarrollo de la evaluación del aprendizaje, y la aplicación de los procedimientos de evaluación previstos (cumplimiento coherente). La primera estaría recogida en la dimensión primera del modelo de competencia evaluativa (*Planificación estratégica*) quedando las dos restantes sujetas a la dimensión cuarta (*Implementación del sistema de evaluación*).
- Rizo (1999) desarrolló en el contexto formativo colombiano un estudio similar al que está siendo presentado, contando con escalas de evaluación de la calidad docente de 53 instituciones diferentes. Su objetivo era detallar los aspectos considerados en la evaluación de la calidad de la enseñanza universitaria. Identificó seis variables principales, una de ellas, la forma de evaluar. Los indicadores sujetos a su desarrollo también se han detectado entre las escalas nacionales y están recogidas en las dimensiones del modelo elaborado: la pertinencia de la evaluación (relacionada con la adecuación pedagógica procurada en la dimensión primera y revisada en la octava del modelo propuesto) y la calificación objetiva (centrada en la precisión recogida en la dimensión quinta y valorada en la octava).

- En el contexto de la creación y validación de un instrumento para la valoración de la competencia docente del profesor universitario (el CEDA), J. M. García (1997) realizó una síntesis de las principales dimensiones implicadas en otras herramientas administradas al alumnado universitario. Su conclusión fue que seis categorías están presentes de manera recurrente, siendo una de ellas la evaluación y calificación del rendimiento discente.
- Con el objetivo de confeccionar un perfil de competencias para el profesorado universitario, I. González y López (2010) y A. B. López, González y León (2014) han presentado recientemente los resultados del análisis de 32 herramientas para la evaluación de la calidad docente, de otras tantas instituciones formativas de todo el estado. El análisis del banco de ítems configurado, les ha llevado a delimitar un total de nueve dimensiones relevantes para la docencia⁷⁵, estando referida una de ellas a los sistemas de evaluación. Dentro de la misma, se incluyen referencias a la variedad de instrumentos evaluativos aplicados, a la información sobre el sistema de evaluación y a su alineación con los contenidos y actividades trabajados. Todos estos aspectos se encuentran representados en la propuesta de sistematización elaborada, principalmente en las dos primeras dimensiones que la conforman.

4. Modelo de competencia evaluativa del profesorado universitario: desarrollo

Los apartados anteriores han introducido las ocho dimensiones que constituyen el modelo de competencia evaluativa docente y han revisado el proceso de génesis que ha conducido a su propuesta. La consideración de los referentes recorridos y, en particular, la teoría de la evaluación detallada en el modelo ITC, avalan su pertinencia y aportan elementos para su justificación.

Este último apartado plantea el desarrollo de los diferentes niveles que jalonan la estructuración del modelo, recorriendo los indicadores y descriptores que conforman cada dimensión. El diseño de los instrumentos de recogida de información empleados para la

⁷⁵ A sus propósitos, competencias. La propuesta de I. González y López (2010) ha sido una de las consideradas para introducir en el capítulo inicial del marco teórico los perfiles profesionales del profesorado en la era de la convergencia europea.

exploración empírica se ha realizado siguiendo esta estructura inicial, para comprobar posteriormente la estructuración interna que revelan las respuestas producidas por la muestra.

Por cada una de las ocho dimensiones se dispone de un subapartado, recorriendo en todos ellos se recorre un mismo itinerario de análisis:

- *Descripción general de la dimensión*, de modo que se establezca un marco común de referencia que favorezca la interpretación unívoca de su formulación. Evitando reiterar la justificación de las condiciones incorporadas en el modelo ITC en el capítulo anterior, en cada caso se incluyen sólo aquellas particularidades que afecten a la competencia que al respecto debe manifestar el profesorado en su actuación.
- *Presentación del sistema de indicadores y de los descriptores de logro previstos en su rúbrica o rejilla de valoración*. Presentada en el **ANEXO 05**, dicha herramienta delimita cinco niveles de logro por cada uno de los indicadores de competencia evaluativa reseñados (inadecuado, poco adecuado, adecuado, muy adecuado y excelente).
- *Análisis de los principales componentes que integran la competencia*. En este apartado se detallan los principales conocimientos básicos, procedimientos, y elementos actitudinales necesarios para el desarrollo y puesta en práctica de cada una de las dimensiones. De acuerdo con las orientaciones ofrecidas por el INTASC (2011) este esquema de análisis debe interpretarse considerando dos reservas:
 - ✓ En primer lugar, la presentación diferenciada de componentes no debe llevar a interpretar una falsa relación de independencia entre los mismos. Muy al contrario, conocimientos, procedimientos y actitudes o predisposiciones intervienen y se movilizan de forma conjunta en cada una de las dimensiones. Su delimitación diferenciada sólo trata de ofrecer una visión amplia del alcance de las mismas.
 - ✓ En segundo lugar, la enumeración de componentes no debe recibirse como una lista de control cerrada. Para favorecer la comprensión de las ocho dimensiones, detalla los principales componentes declarativos, procedimentales y actitudinales que se asocian a su manifestación práctica, pero ni son los únicos, ni son indispensables en su totalidad. Admitida la versatilidad y el carácter reconstructivo como rasgos propios de las competencias en general (Cano, 2007, 2011; Le Boterf, 2001; Unigarro &

Rondón, 2005), toda delimitación de componentes debe interpretarse en términos de flexibilidad y apertura.

4.1 Planificación estratégica de la evaluación para el aprendizaje.

Planificar estratégicamente la evaluación del aprendizaje para promover en las/os estudiantes el logro de las competencias previstas en el currículo y su acreditación, a partir de una adecuada cantidad y variedad de evidencias recabadas a lo largo de todo el proceso formativo.

a) Descripción de la dimensión.

Esta primera dimensión refiere de forma directa a la fase de planificación de la evaluación, concibiéndola como un proceso inherente a la enseñanza y el aprendizaje. Conecta directamente con el subproceso de planificación estratégica del modelo ITC (**ANEXO 04**) e incorpora en su desarrollo las exigencias que derivan de cuatro de los requisitos que lo integran, todos salvo los referidos a la evaluación participativa y a la atención a la diversidad, que pasan a considerarse en las dimensiones séptima y tercera, respectivamente.

Esta primera dimensión presenta la evaluación como un recurso profesional de gran potencialidad didáctico-pedagógica por su incidencia en el trabajo y el esfuerzo de docentes y discentes, y por los correspondientes efectos que ello comporta en el aprendizaje. De modo similar, Toledo (2005) relaciona la competencia evaluativa del profesorado con su capacidad de concebir y generar situaciones evaluativas acordes al nivel del alumnado y con una visión amplia de los objetivos que deben alcanzarse.

El empleo estratégico de la evaluación como recurso didáctico hace de ella un complejo proceso de elección y toma de decisión que debería ser consciente e informado (Brown, 2003b; J. Chappuis, 2014; McMillan, 2000). En concreto, la delimitación de esta primera dimensión identifica cuatro espacios de decisión vinculados a la planificación evaluativa, trasladados a los cuatro indicadores expuestos en el siguiente apartado. Se reconoce así que el profesorado competente en este terreno, es aquel que idea conscientemente sistemas de evaluación que promuevan el aprendizaje y no sólo su calificación. Para ello, formula demandas acordes con las competencias a desarrollar y con el trabajo previsto en el aula, intercala y gradúa estratégicamente los usos pedagógicos y acreditativos a lo largo

del proceso formativo y elige las herramientas y procedimientos que mejor se ajusten a sus propósitos y a las condiciones prácticas.

La necesidad e importancia de la planificación evaluativa responde a una de las premisas fundamentales del enfoque teórico de la evaluación para el aprendizaje y así ha sido justificado al presentar el modelo ITC. Su pertinencia como dimensión destacada dentro de la competencia evaluativa docente se ve reforzada por el reconocimiento unánime que encuentra en los antecedentes que han tratado de definirla en distintos contextos formativos (formación básica de maestras/os, docencia universitaria o enseñanza en general) (tabla 46). Su relevancia y visibilidad también se advierte al analizar el contenido de las escalas empleadas en la evaluación de la calidad docente y en los cuestionarios aplicados en estudios sobre la praxis evaluativa en la universidad. Sin hacer referencia expresa a la planificación (figura 24), en ambos casos son frecuentes las referencias a cuestiones como la continuidad, la orientación formativa de la evaluación, la alineación o la combinación de instrumentos diferentes.

b) Indicadores clave en su desarrollo.

El desarrollo de esta primera competencia exige centrar la atención en cuatro indicadores prioritariamente. Según lo dispuesto en la rúbrica de referencia (**ANEXO 05**), por cada uno de ellos se han definido descriptores para cinco niveles de logro ordenados de menor a mayor cumplimiento.

- *Articula un sistema de evaluación continua que, sensible al proceso de construcción que efectúan las/os estudiantes en su aprendizaje, recabe evidencias en las distintas fases del mismo.*

INADECUADO	PREDOMINIO DE LA EVALUACIÓN FINAL. Las tareas evaluativas que plantea a las/os estudiantes se concentran básicamente en una determinada fase del proceso formativo (por lo general, en el tramo final).
POCO ADECUADO	EVALUACIÓN FINAL Y DEL PROCESO. Las tareas evaluativas que plantea al alumnado se concentran fundamentalmente en dos de las fases del proceso formativo (por lo general, en las fases central y final).
ADECUADO	EVALUACIÓN CONTINUA NO ATOMIZADA. Propone tareas evaluativas al comienzo del proceso formativo, durante su desarrollo y al finalizarlo, cuidando no caer en la atomización de la valoración del aprendizaje.
MUY ADECUADO	EVALUACIÓN CONTINUA DE INTENSIDAD GRADUADA. Propone tareas evaluativas en las distintas fases del ciclo formativo, determinando la concentración e intensidad de las mismas en función del proceso de aprendizaje que deben llevar a cabo las/os estudiantes (momentos críticos, tiempos de asimilación, etc.).

EXCELENTE	EVALUACIÓN CONTINUA GRADUADA Y COORDINADA. Propone tareas evaluativas en las distintas fases del ciclo formativo, determinando la concentración e intensidad de las mismas en función del proceso de aprendizaje que debe realizar el alumnado y en coordinación con otras asignaturas (previsión global de la curva de esfuerzo, distribución equilibrada de carga de trabajo).

- *Plantea una propuesta evaluativa alineada con los objetivos formativos y las estrategias metodológicas previstas para su consecución.*

INADECUADO	DISTANCIA ENTRE EVALUACIÓN y OBJETIVOS. Las actividades que plantea para evaluar el aprendizaje de las/os estudiantes no responden con fidelidad a las competencias detalladas en el currículo.
POCO ADECUADO	DISTANCIA ENTRE EVALUACIÓN y METODOLOGÍA. Propone actividades diferentes para desarrollar las competencias detalladas en el currículo y para realizar su evaluación.
ADECUADO	ALINEACIÓN O COHERENCIA. Las actividades evaluativas que plantea valoran las competencias previstas en el currículo y están directamente relacionadas con el trabajo que desarrollamos en el día a día.
MUY ADECUADO	ALINEACIÓN PLENA. En el día a día plantea a las/os estudiantes el mismo tipo de actividades que servirán para la evaluación del grado de consecución de las competencias previstas.
EXCELENTE	INTEGRACIÓN. La única diferencia entre las actividades de aprendizaje y de evaluación planteadas es el objetivo prioritario con el que se desarrollan: ambas ayudan a las/os estudiantes a comprender lo que se espera de ellas/os y sirven lo mismo para su desarrollo que para su evaluación.

- *Diseña un sistema evaluativo que combine sus funciones pedagógicas y acreditativas para orientar estratégicamente a las/os estudiantes en el desarrollo de las competencias previstas en el currículo.*

INADECUADO	EVALUACIÓN ACREDITATIVA. La mayor parte de las actividades evaluativas que plantea durante el curso cuentan para nota.
POCO ADECUADO	PREDOMINIO ACREDITATIVO. Combina actividades evaluativas destinadas a mejorar el aprendizaje con otras destinadas a determinar la calificación, pero no de forma sistemática ni proporcionada (la mayor parte de las tareas evaluativas cuentan para nota).
ADECUADO	COMBINACIÓN ESTRATÉGICA. Combina de forma consciente y deliberada tareas de evaluación acreditativa y pedagógica (de mejora) para lograr el desarrollo de las competencias previstas, no sólo su medición.
MUY ADECUADO	COMBINACIÓN ESTRATÉGICA: REGULACIÓN DEL TIEMPO y el ESFUERZO. Idea un sistema de evaluación que lleva a las/os estudiantes a dedicar el tiempo y esfuerzo que requiere su aprendizaje, mediante la combinación deliberada de tareas pedagógicas y acreditativas.
EXCELENTE	COMBINACIÓN ESTRATÉGICA: REGULACIÓN DEL TIEMPO y el ESFUERZO y su GESTIÓN. Idea un sistema de evaluación que lleva a las/os estudiantes a dedicar el tiempo y esfuerzo que requiere su aprendizaje y a gestionarlo de manera efectiva a lo largo del proceso formativo, mediante la combinación deliberada de tareas pedagógicas y acreditativas.

- *Prevé el uso de instrumentos y procedimientos de evaluación pertinentes, variados y adecuados al contexto evaluativo específico para obtener evidencias sobre el aprendizaje de las competencias.*

INADECUADO	CENTRADA EN EL CONTENIDO. Emplea técnicas y procedimientos que atienden más al dominio de contenidos (reproducción, descripción teórica, aplicación descontextualizada) que al desarrollo de competencias (aplicación en situaciones reales o simuladas).
POCO ADECUADO	INADECUACIÓN PARA LA EVALUACIÓN DE COMPETENCIAS. Prevé el uso de procedimientos e instrumentos de evaluación para evaluar las competencias previstas en el currículo y no meros contenidos, pero detecta imprecisiones en las valoraciones resultantes y/o desajustes de carácter práctico u organizativo (interferencias con la normativa académica, sobrecarga, etc.).
ADECUADO	VARIADA y ADECUADA al CONTEXTO. Prevé el uso de procedimientos e instrumentos de evaluación variados, capaces de proporcionar evidencias sobre el grado de desarrollo de las competencias en las/os estudiantes y aplicables en las condiciones prácticas previstas para su desarrollo (normativa, número de estudiantes, plazos, etc.).
MUY ADECUADO	PRECISA, VARIADA y ADECUADA al CONTEXTO. Prepara un repertorio variado de instrumentos y procedimientos de evaluación, capaz de captar con precisión distintos niveles de desarrollo en las competencias y aplicable en las condiciones reales de trabajo previstas (número de estudiantes, normativa, académica, carga de trabajo, etc.).
EXCELENTE	PRÁCTICA, VARIADA, PRECISA, VIABLE y ADECUADA AL GRUPO. Plantea situaciones evaluativas muy próximas a la realidad, combinando instrumentos y procedimientos precisos, viables, y adecuados a los intereses y necesidades del grupo de estudiantes.

c) *Componentes integrantes.*

Movilizados de forma armónica y conjunta según precise la complejidad de cada situación particular, algunos de los elementos más representativos que integran esta dimensión de la competencia evaluativa docente son los contemplados en la tabla 49. Según las precisiones efectuadas al comienzo del apartado, el contenido de las tres columnas no es independiente ni determinante, de modo que caben distintas manifestaciones en su puesta en práctica:

Tabla 49

Principales componentes integrados en la planificación estratégica de la evaluación para el aprendizaje (dimensión 01)

PLANIFICACIÓN ESTRATÉGICA DE LA EVALUACIÓN PARA EL APRENDIZAJE		
BASE CONCEPTUAL Conocimientos, saberes, dominio de...	BASE PROCEDIMENTAL Capacidades en acción, desarrollo práctico	BASE ACTITUDINAL Actitudes, predisposiciones, principios, inquietudes
• Proceso de evaluación (fases, funciones, normas, tipos, etc.). • Proceso de aprendizaje en personas adultas y principios didácticos relacionados. • Dificultades generales y tópicos del aprendizaje disciplinar más problemáticos. • Perfil profesional de referencia y distribución de competencias en el plan de estudios. • Rasgos e implicaciones básicas de la evaluación de competencias. • Principales efectos de la evaluación sobre el aprendizaje y condiciones que lo favorecen. • Modelos de planificación evaluativa. • Normativa en materia evaluativa. • Repertorio de instrumentos y técnicas de evaluación, y sus ventajas y limitaciones asociadas. • Principales efectos de los distintos instrumentos de evaluación sobre el esfuerzo y aprendizaje discente. • Conceptos básicos vinculados a la validez y fiabilidad de las herramientas de recogida de evidencias.	• Diseño de experiencias de aprendizaje significativo en el contexto didáctico específico de la formación universitaria. • Planificación justificada de procesos de evaluación del aprendizaje. • Selección de evidencias significativas de los aprendizajes previstos. • Combinación equilibrada y razonada de los usos pedagógicos y acreditativos de la evaluación. • Anticipación de los efectos principales de las decisiones evaluativas adoptadas en el aprendizaje y la enseñanza. • Diseño de actividades evaluativas para la recogida de evidencias significativas sobre el desarrollo de las competencias previstas. • Justificación pedagógica de las herramientas de evaluación seleccionadas. • Combinación equilibrada y razonada de instrumentos de evaluación de diferente tipo. • Identificación realista de lagunas y limitaciones en las herramientas seleccionadas o diseñadas. • Previsión ajustada de la carga de trabajo docente y discente que generan los procedimientos evaluativos. • Graduación coordinada y sensible de tareas y demandas. • Anticipación de los efectos de los procedimientos de evaluación seleccionados en el aprendizaje. • Adecuación de procedimientos evaluativos a las circunstancias particulares. • Valoración objetiva de la validez y fiabilidad de instrumentos aplicables.	• Orientación al aprendizaje. • Expectativas positivas y confianza en las posibilidades de aprendizaje del alumnado. • Compromiso con el perfil profesional y el desarrollo colegiado del plan de estudios. • Rigor y precisión. • Elección razonada. • Sistematicidad, estrategia como principio de actuación. • Realismo. • Creatividad e ingenio. • Actualización permanente e innovación.

4.2 Explicitación y transparencia.

Clarificar con las/os estudiantes los criterios y condiciones que rigen la evaluación del aprendizaje para establecer una visión clara y unívoca de la demostración de los logros a alcanzar que oriente su trabajo a lo largo del proceso.

a) Descripción de la competencia.

Abordada expresamente en el segundo subproceso del modelo ITC, la explicitación y transparencia en los criterios y condiciones de la evaluación del aprendizaje representa un requisito vinculado a su calidad de forma unánime y generalizada. Según lo detallado en el capítulo anterior, su necesidad y pertinencia se justifica desde diferentes perspectivas: desde el punto de vista legal o normativo, como derecho del alumnado y obligación de las/os responsables de la evaluación, y también desde el pedagógico, como factor que contribuye a la clarificación de las expectativas de aprendizaje y la regulación del esfuerzo y el trabajo hacia las mismas.

La revisión de los referentes para la sistematización de la competencia evaluativa docente conduce a esa misma conclusión. La información clara y explícita sobre las premisas y condiciones de la evaluación es el tópico evaluativo más reiterado en las escalas institucionales para la evaluación de la calidad docente (figura 24) y uno de los más representados para valorar las prácticas evaluativas universitarias. La eficacia del profesorado en el establecimiento de criterios y condiciones públicas y transparentes en el diseño y puesta en práctica del proceso de evaluación también aparece contemplada en las propuestas de sistematización previas de Toledo (2005, 2006), Stiggins (1998) o Monereo y Castelló (2009a). Otro dato que avala su relevancia es que para el alumnado representa uno de los rasgos que caracteriza al profesorado universitario mejor valorado (V. Álvarez et al., 1999; Bain, 2007; B. Gargallo et al., 2010).

En consecuencia, parece fundamental que el profesorado muestre solvencia haciendo conocidos y asequibles a las/os estudiantes los términos de la evaluación de su aprendizaje. De acuerdo con lo establecido en el modelo ITC, para favorecerlo las/os docentes deberían trabajar a niveles complementarios, garantizando la explicitación de criterios y compromisos evaluativos públicos primero, y su comprensión e interiorización por parte del alumnado después. Ello exige el desarrollo de acciones y actuaciones específicas a lo largo del proceso formativo, no sólo al comienzo. Algunas de ellas pueden tener lugar por medio de

los procesos de retroalimentación y participación discente abordados en próximas dimensiones.

b) Indicadores clave en su desarrollo.

El desarrollo de la segunda dimensión se concreta en un único indicador. Para el mismo se proponen cinco niveles de logro como en el resto de casos (**ANEXO 05**):

- *Desarrolla actividades específicas para facilitar la interiorización y apropiación de los criterios de evaluación por parte de las/os estudiantes a lo largo del proceso formativo.*

INADECUADO	EXPLICITACIÓN. Detalla por escrito los criterios y condiciones generales que rigen el proceso de evaluación del aprendizaje en la documentación básica de la asignatura// No hay guía de aprendizaje, no se facilita la documentación.
POCO ADECUADO	EXPLICITACIÓN y PRESENTACIÓN. Al comienzo del curso dedica un tiempo a exponer y explicar los criterios y condiciones generales de evaluación detallados por escrito en la documentación de la asignatura.
ADECUADO	EXPLICITACIÓN, PRESENTACIÓN y CLARIFICACIÓN. Durante el curso lleva a cabo actividades para clarificar el significado de los criterios de evaluación explicitados, y lograr que las/os estudiantes comprendan realmente los logros hacia los que deben encauzar su trabajo (recordatorios, ejemplos, aclaración de significados, etc.).
MUY ADECUADO	HERRAMIENTAS DE CLARIFICACIÓN. Partiendo de los criterios generales de evaluación explicitados, presenta a las/os estudiantes herramientas o materiales de apoyo que clarifiquen su concreción y transferencia a la valoración de tareas, unidades o actividades concretas (rúbricas, listas de control, modelos, etc.).
EXCELENTE	APLICACIÓN HERRAMIENTAS DE CLARIFICACIÓN. Propone actividades para que las/os estudiantes trabajen con las herramientas de valoración creadas (rúbricas, listas de control, etc.) y las apliquen en situaciones de autoevaluación o evaluación de compañeros para unificar interpretaciones y profundizar en su comprensión.

Las recomendaciones recopiladas en la fase de consulta a expertas/os sugerían simplificar el modelo de competencia evaluativa para hacerlo manejable. Ello ha llevado a proponer dos dimensiones compuestas de un único indicador: esta segunda, referida a la clarificación de los criterios y condiciones de la evaluación y la séptima, centrada en la participación discente.

Como será expuesto al justificar el diseño del marco empírico de la investigación, esta medida supone la asunción de ciertos riesgos metodológicos con especial incidencia en la validez o estructuración interna de las respuestas. Su comprobación se realiza agrupando los ítems/indicadores con correlaciones moderadas y altas entre sí e identificando los factores resultantes (Mateo, 2009b). Un factor debe estar

conformado al menos por dos ítems o indicadores, y preferiblemente por tres o más (Beavers et al., 2013; Costello & Osborne, 2005; Morales, 2013). Cuando la dimensión teórica está formada por un único elemento, la exploración de la estructura interna forzará su relación con otros indicadores, rompiendo con la dimensionalidad prevista inicialmente.

Por consiguiente, la consecuencia previsible de la decisión adoptada es una estructuración interna diferente a la que está siendo presentada, con repercusión potencial en la consistencia interna (fiabilidad) de los factores resultantes. En contrapartida, las razones para asumir dicho riesgo y mantener la configuración de un único componente son dos principalmente:

- ✓ La primera de ellas es que la definición de cinco niveles de logro permite captar los distintos requisitos vinculados a esta dimensión en los referentes consultados. Cubierta esa premisa, la simplificación del modelo se pone por delante de la posibilidad de mejorar su estructura interna incorporando elementos de refuerzo.
- ✓ La segunda es que la estructura presentada permite explorar desde la perspectiva empírica las relaciones de esta dimensión de la competencia evaluativa con otras. De hecho, no existe para la misma una única alternativa de agrupación; la clarificación de criterios y condiciones evaluativas guarda relación con varios procesos asociados a la puesta en práctica de la evaluación (de ajuste, retroalimentación o participación, por ejemplo), pero también es consecuencia de una planificación rigurosa.

c) Componentes integrantes.

Movilizados de forma armónica y conjunta según precise la complejidad de cada situación particular, algunos de los elementos más representativos que integran esta dimensión de la competencia evaluativa docente son los contemplados en la tabla 50:

Tabla 50

Principales componentes integrados en la explicitación y transparencia de la evaluación (dimensión 02)

EXPLICITACIÓN Y TRANSPARENCIA		
BASE CONCEPTUAL Conocimientos, saberes, dominio de...	BASE PROCEDIMENTAL Capacidades en acción, desarrollo práctico	BASE ACTITUDINAL Actitudes, predisposiciones, principios, inquietudes
- Efecto de las percepciones en la configuración de representaciones educativas profundas. - Normativa en materia evaluativa. - Representaciones evaluativas prototípicas en el alumnado universitario: prejuicios y preconceptos. - Dificultades típicas en la interpretación de los estándares y criterios profesionales de calidad. - Repertorio de actividades y estrategias para el ajuste de representaciones.	- Comunicación clara y adecuada de los criterios y condiciones evaluativas en formato escrito. - Presentación oral clara y adecuada de las condiciones evaluativas, al comienzo del proceso, y durante su desarrollo. - Identificación de las representaciones iniciales del alumnado sobre el aprendizaje a efectuar y su evaluación. - Detección de interpretaciones erróneas en los criterios y estándares profesionales de calidad. - Confrontación de las representaciones, concepciones o expectativas erróneas del alumnado. Selección de ejemplos y modelos significativos para la clarificación y ajuste de las representaciones discentes. - Diseño de herramientas y situaciones para la clarificación y ajuste de las representaciones discentes.	- Empatía. - Diálogo, comunicación. - Persuasión. - Transparencia, claridad. - Cautela. - Perseverancia.

4.3 Diversificación de la evaluación.

Adaptar las condiciones sobre las que se realiza la evaluación del aprendizaje para responder a la diversidad de necesidades de las/os estudiantes, considerando los criterios mínimos definidos de forma general y preservando la transparencia y justicia del proceso evaluativo.

a) Descripción de la dimensión.

Esta tercera dimensión se centra en la respuesta que articula el profesorado desde el proceso de evaluación para atender a la diversidad natural de estilos y

necesidades de aprendizaje de las/os estudiantes. Por medio de decisiones y actuaciones que pueden tener lugar en la fase de planificación y/o de implementación, apunta a la configuración de dispositivos evaluativos no estereotipados, que detecten e identifiquen las necesidades particulares del alumnado en el proceso formativo y ofrezcan una respuesta ajustada a las mismas.

Por su conexión con los referentes teóricos de la evaluación para el aprendizaje y la evaluación de competencias, el modelo ITC establece que la diversificación representa un rasgo asociado a la calidad evaluativa. Sin embargo, destaca como la dimensión más controvertida y menos secundada de todas las incorporadas en la propuesta de sistematización de la competencia docente. Como ya ha sido indicado en el capítulo cuarto, los referentes institucionales y normativos para su ejercicio se orientan más a la uniformidad y la homogeneidad que a la divergencia evaluativa. Así se constata también al revisar el contenido de las escalas de evaluación de la calidad docente y en las herramientas empleadas en la investigación sobre prácticas evaluativas, donde el reflejo del tratamiento a la diversidad resulta anecdótico. La falta de consenso que suscita la diversificación de los dispositivos evaluativos también se ha puesto de manifiesto en la consulta planteada a expertas/os en la materia, con testimonios tanto a favor, como en contra.

Según D. Salinas (2002) las discrepancias en el trasfondo de esta dimensión están asociadas a los dilemas que genera su repercusión acreditativa. La diversificación de los mecanismos evaluativos aparece ampliamente suscrita como medida de apoyo al aprendizaje (Ahumada, 2002, 2005; Cano et al., 2010; V. M. López et al., 2007; Santos, 2006), aunque rara vez se contemplan las consecuencias de este tipo de medidas en el marco de una evaluación criterial y sumativa. El principio de la justicia de la evaluación formulado por Baker et al. (1996) reconoce el dilema y trata de ponerle remedio conjugando dos condiciones en los procesos evaluativos: el establecimiento de medidas en sentido compensatorio para favorecer que cada estudiante pueda desarrollar y demostrar aprendizaje desde su situación particular y la garantía de que en los procesos de puntuación y calificación no se produzca ningún trato de favor. Este equilibrio adopta una especial significación en el contexto específico de las profesiones reguladas; según se desprende del modelo ITC el profesorado debería prever diferentes alternativas evaluativas para acreditar los mismos aprendizajes, aquellos señalados en los decretos reguladores y redefinidos en los perfiles profesionales de los planes formativos.

En consecuencia, la tercera dimensión plantea que el profesorado debe introducir ajustes y medidas para adecuar las condiciones de la evaluación a la diversidad del

alumnado, preservando la aplicación transparente de los criterios que determinan la valoración del aprendizaje. Para ello, debe partir necesariamente de la identificación de la situación de partida de las/os estudiantes con respecto al aprendizaje a efectuar (Monereo & Castelló, 2009a), para favorecer la adopción de las medidas de ajuste necesarias y su seguimiento a lo largo del proceso (Quesada et al., 2013).

b) Indicadores clave en su desarrollo.

En coherencia con la descripción efectuada, la tercera dimensión se desarrolla por medio de dos indicadores de evaluación complementarios, y sus respectivos descriptores de logro o realización (**ANEXO 05**):

- *Determina el nivel de partida de las/os estudiantes con respecto al aprendizaje a efectuar, identificando aspectos que puedan condicionar de forma relevante la evolución y desarrollo del proceso.*

INADECUADO	SIN EVALUACIÓN INICIAL. No lleva a cabo actividades de evaluación inicial que determinen el punto de partida de las/os estudiantes con respecto al aprendizaje a realizar.
POCO ADECUADO	EVALUACIÓN INICIAL NO SISTEMÁTICA. Al comienzo del proceso lleva a cabo actividades para determinar el punto de partida del grupo de estudiantes con respecto al aprendizaje a realizar, pero no siempre ni de forma rigurosa.
ADECUADO	EVALUACIÓN INICIAL. Al comienzo del proceso lleva a cabo actividades para identificar y valorar deliberadamente el nivel de partida de las/os estudiantes con respecto al aprendizaje a realizar y sus necesidades particulares.
MUY ADECUADO	EVALUACIÓN INICIAL y SEGUIMIENTO. Realiza un seguimiento sistemático del grupo de estudiantes que permite valorar su evolución y sus necesidades a lo largo del proceso formativo.
EXCELENTE	EVALUACIÓN INICIAL y SEGUIMIENTO INDIVIDUALIZADO. Realiza un seguimiento individualizado de las/os estudiantes, que permite valorar la evolución de cada una/o de ellas/os y sus necesidades particulares a lo largo del proceso formativo.

- *Introduce ajustes y medidas para adecuar las condiciones de la evaluación a la diversidad del alumnado garantizando la transparencia del proceso.*

INADECUADO	EVALUACIÓN UNIFORME. Los criterios y condiciones que establece para la evaluación del aprendizaje son iguales y comunes para todas/os las/os estudiantes.
POCO ADECUADO	FLEXIBILIZACIÓN PUNTUAL. De forma excepcional, flexibiliza las condiciones de la evaluación para responder a las necesidades particulares de determinadas/os estudiantes.
ADECUADO	DIVERSIFICACIÓN DE CONDICIONES EVALUATIVAS. Introduce adaptaciones en las condiciones de la evaluación, de acuerdo a pautas de diversificación públicamente establecidas (quién, cómo y cuándo puede acceder a tales adecuaciones).

MUY ADECUADO	VÍAS ALTERNATIVAS. Ofrece a las/os estudiantes vías alternativas para la evaluación, en el marco de las condiciones definidas para la diversificación y de acuerdo a los criterios generales de logro establecidos.
EXCELENTE	NEGOCIACIÓN PERSONALIZADA. Negocia con las/os estudiantes alternativas evaluativas personalizadas para responder a necesidades particulares, respetando las condiciones definidas para la diversificación y los criterios generales de logro establecidos.

c) Componentes integrantes.

Movilizados de forma armónica y conjunta según precise la complejidad de cada situación particular, algunos de los elementos más representativos que integran esta dimensión de la competencia evaluativa docente son los contemplados en la tabla 51:

Tabla 51
Principales componentes integrados en la diversificación de la evaluación (dimensión 03)

DIVERSIFICACIÓN DE LA EVALUACIÓN		
BASE CONCEPTUAL Conocimientos, saberes, dominio de...	BASE PROCEDIMENTAL Capacidades en acción, desarrollo práctico	BASE ACTITUDINAL Actitudes, predisposiciones, principios, inquietudes
- Perfil general del grupo de estudiantes y rasgos particulares de las/os integrantes. - Medidas y principios didáctico-pedagógicos para la atención a la diversidad. - Procedimientos de evaluación diagnóstica o inicial. - Perfil profesional de referencia y criterios y/o estándares delimitados para su evaluación en el plan de estudios - Incidencia de la ideología y los valores socioculturales en la evaluación. - Repertorio de instrumentos y técnicas de evaluación, y sus ventajas y limitaciones asociadas. - Normativa evaluativa de referencia.	- Identificación del nivel de partida del alumnado con respecto a los aprendizajes a realizar y de sus necesidades particulares de apoyo. - Formulación de medidas para diversificar las condiciones de la evaluación sin perjuicio de la justicia y la transparencia del proceso de calificación (plazos flexibles o diversos, optatividad, combinación de instrumentos, tareas adecuadas, etc.). - Adecuación del dispositivo evaluativo inicial a la diversidad natural presente en el grupo. - Control del riesgo de agravio comparativo. - Diseño de itinerarios evaluativos diferentes para la valoración de los mismos aprendizajes. - Negociación de acuerdos para la adecuación de las condiciones evaluativas. - Documentación de los acuerdos adoptados en materia evaluativa.	- Orientación al aprendizaje. - Respeto a la diversidad. - Expectativas positivas y confianza en las posibilidades de aprendizaje del alumnado. - Creatividad e ingenio. - Flexibilidad. - Transparencia. - Equidad, no discriminación. - Comunicación, diálogo, negociación.

4.4 Implementación ajustada.

Implementar los sistemas de evaluación planificados para garantizar el cumplimiento de sus propósitos, introduciendo y documentando los ajustes necesarios para responder a las condiciones prácticas reales y a los imprevistos surgidos.

a) Descripción de la dimensión.

Esta nueva dimensión de la competencia evaluativa tiene lugar en la fase de implementación o puesta en práctica del proceso de evaluación del aprendizaje y se corresponde con el primer criterio del subproceso de recogida de evidencias del modelo ITC. Hace referencia a la capacidad del profesorado universitario para dirigir la evaluación del aprendizaje de acuerdo a dos referencias indispensables: la previsión inicial efectuada para la misma antes de comenzar el proceso y la situación evidenciada en el contexto real de aplicación durante su transcurso. Compaginadas de forma armónica y consciente, el desarrollo coherente de lo dispuesto en la planificación evaluativa y la introducción de los ajustes y modificaciones pertinentes representan condiciones igualmente necesarias, para consolidar prácticas evaluativas de calidad.

El desarrollo coherente del dispositivo evaluativo diseñado ha sido presentado en el capítulo cuarto como una condición unánime en los referentes normativos para la praxis evaluativa. Las escalas de evaluación de la calidad de la docencia administradas al alumnado, también inciden en este indicador especialmente, siendo uno de los temas más presentes junto a la transparencia y los aspectos asociados a la planificación (figura 24). Su relevancia práctica se reconoce también desde algunas de las propuestas que han sistematizado la competencia evaluativa para la enseñanza en niveles escolares (American Association of School Administrators, National Association of Elementary School Principals, National Association of Secondary School Principals, & National Council on Measurement in Education, 1997; American Federation of Teachers, National Council on Measurement in Education, & National Education Association, 1990).

La introducción de ajustes en las previsiones iniciales y la gestión de los cambios resultantes tiene una representación mucho menor entre los referentes consultados. La flexibilización de la planificación evaluativa, a menudo es percibida como una falta o incumplimiento por parte del propio profesorado (Serrano, 2010). Su presencia entre los indicadores del modelo ITC se debe a la adopción de las competencias como objeto evaluativo en el EEES, que incrementa sustancialmente

la complejidad del proceso. Compartida tal premisa, Monereo y Castelló (2009a) también reconocen en su propuesta que el profesorado competente en materia evaluativa debe prever márgenes de flexibilidad para adecuar los dispositivos creados a las exigencias de las condiciones prácticas reales.

En este contexto, empieza a cobrar relevancia la documentación y gestión efectiva de los cambios resultantes en la evaluación, que contempla su comunicación, su justificación y, en la medida de lo posible, el consenso con el alumnado. Este requisito cobra una especial significación en la formación básica de maestras/os, ya que aproxima a las/os futuras/os evaluadoras/es a la complejidad evaluativa y al juicio profesional asociado a ella (I. Álvarez, 2008a; V. M. López, Castejón, et al., 2012; Shepard, 2000; Ureña et al., 2009).

Los indicadores presentados a continuación hacen referencia a este balance e incorporan respectivamente el desarrollo coherente de la planificación evaluativa y la gestión de las adecuaciones realizadas a la misma durante la puesta en práctica.

b) Indicadores clave en su desarrollo.

Como en el resto de casos, la rúbrica de referencia (**ANEXO 05**) detalla cinco niveles de logro para cada uno de los dos indicadores incluidos en esta cuarta dimensión del modelo:

- *Pone en práctica el sistema evaluativo de acuerdo a las decisiones adoptadas en la fase de planificación, atendiendo a las principales condiciones que median en su aplicación.*

INADECUADO	APLICACIÓN DIRECTA. Aplica el sistema de evaluación tal y como ha sido diseñado, sin tener en cuenta condiciones o sucesos imprevistos que surjan durante su puesta en práctica.
POCO ADECUADO	ADECUACIONES PUNTUALES. Trata de adecuar el sistema evaluativo diseñado a las condiciones prácticas que determinan su desarrollo, pero no siempre ni de forma sistemática (atención a imprevistos, ajustes puntuales de calendario o espacios, etc.).
ADECUADO	CONDICIONES TÉCNICAS. Pone en práctica el sistema evaluativo diseñado atendiendo a las principales condiciones que pueden incidir en su desarrollo desde el punto de vista técnico (amenazas a la legitimidad, ventajas o desventajas vinculadas al orden de realización o a plazos de entrega diferentes, etc.).
MUY ADECUADO	CONDICIONES TÉCNICAS y RELACIONALES. Pone en práctica el sistema evaluativo diseñado atendiendo a las condiciones que pueden incidir en su desarrollo tanto desde el punto de vista técnico, como del relacional (clima de tensión o distensión excesiva, condiciones físicas desfavorables, etc.).
EXCELENTE	APLICACIÓN CONTROLADA. Aplica el sistema evaluativo diseñado adoptando medidas para favorecer las condiciones técnicas y relacionales idóneas para su correcto desarrollo, y valorando el impacto potencial de los imprevistos y factores no controlados.

- *Documenta de forma justificada los ajustes y modificaciones introducidos en el desarrollo del proceso evaluativo en respuesta a los imperativos prácticos e imprevistos.*

INADECUADO	APLICACIÓN LIBRE. Aplica el sistema evaluativo con total flexibilidad, respondiendo sobre el terreno a los incidentes e imprevistos surgidos en su desarrollo.
POCO ADECUADO	SEGUIMIENTO INFORMAL. Lleva a cabo un seguimiento informal de la aplicación del sistema evaluativo, identificando las desviaciones producidas y los ajustes realizados durante su desarrollo (pero sin documentarlas).
ADECUADO	SEGUIMIENTO y DOCUMENTACIÓN DE AJUSTES. Realiza seguimiento del proceso de aplicación del sistema evaluativo, documentando y justificando las desviaciones constatadas y los ajustes efectuados.
MUY ADECUADO	SEGUIMIENTO, DOCUMENTACIÓN y JUSTIFICACIÓN DE AJUSTES. Realiza seguimiento del proceso de aplicación del sistema evaluativo, documentando y justificando ante las/os estudiantes las desviaciones constatadas y los ajustes efectuados.
EXCELENTE	...y CONSENSO DE LOS AJUSTES. Realiza seguimiento del proceso de aplicación del sistema evaluativo, consensuando con las/os estudiantes los ajustes pertinentes y documentando los mismos de forma justificada.

c) Componentes integrantes.

Movilizados de forma armónica y conjunta según precise la complejidad de cada situación particular, algunos de los elementos más representativos que integran esta dimensión de la competencia evaluativa docente son los presentados en la tabla 52:

Tabla 52
Principales componentes integrados en la implementación ajustada de la evaluación (dimensión 04)

IMPLEMENTACIÓN AJUSTADA		
BASE CONCEPTUAL Conocimientos, saberes, dominio de...	BASE PROCEDIMENTAL Capacidades en acción, desarrollo práctico	BASE ACTITUDINAL Actitudes, predisposiciones, principios, inquietudes
• Técnicas y modelos de registro de incidentes. • Modelos flexibles de planificación evaluativa. • Normativa evaluativa de referencia. • Incidencia potencial de las condiciones técnicas y relacionales en la evaluación. • Medidas para favorecer condiciones técnicas y relacionales propicias.	• Previsión de márgenes y espacios de adecuación en la planificación didáctica y evaluativa. • Identificación de las circunstancias prácticas que obstaculizan el normal desarrollo del dispositivo evaluativo diseñado. • Control del efecto potencial de los factores técnicos y ambientales en la evaluación. • Control de imprevistos.	• Compromiso con lo planificado. • Flexibilidad. • Creatividad e ingenio. • Capacidad de regulación, rectificación o ajuste. • Negociación, diálogo, consenso. • Orientación a la mejora. • Transparencia. • Control, observación minuciosa.

-
- Registro de incidencias.
 - Introducción de ajustes y modificaciones sobre la planificación evaluativa inicial.
 - Comunicación razonada de cambios y ajustes.
 - Negociación y discusión de condiciones evaluativas con el alumnado.
 - Propuesta de medidas para responder a imprevistos y circunstancias prácticas.
 - Documentación de ajustes y modificaciones introducidos con respecto a la previsión inicial.
 - Supervisión del cumplimiento en curso de la planificación ideada.
-

4.5 Interpretación y análisis de evidencias.

Analizar e interpretar las evidencias de aprendizaje recopiladas para generar valoraciones válidas y precisas del proceso de aprendizaje y sus resultados, aplicando medidas que contrarresten las principales causas de sesgo e influencia.

a) Descripción de la dimensión.

La quinta dimensión de la competencia evaluativa docente se centra primordialmente a las fases segunda y tercera del proceso evaluativo. Responde al subproceso de análisis de resultados previsto en el modelo ITC y también incorpora la comunicación formal de conclusiones incluido en el subproceso relativo al feedback y la proacción.

Según ha sido justificado en el capítulo anterior, la validez y la fiabilidad representan atributos técnicos de la evaluación admitidos de forma generalizada en referentes teóricos, normativos e institucionales. Así se contempla también en los antecedentes que se han aproximado a la delimitación de la competencia evaluativa docente (American Association of School Administrators, National Association of Elementary School Principals, National Association of Secondary School Principals, & National Council on Measurement in Education, 1997; American Federation of Teachers, National Council on Measurement in Education, & National Education Association, 1990; Monereo & Castelló, 2009a; Stiggins, 1998; Toledo, 2005, 2006).

Las alusiones a la validez de la evaluación y su fiabilidad son frecuentes en las herramientas aplicadas en la investigación de las prácticas evaluativas, así como en las escalas institucionales destinadas a valorar la calidad de la enseñanza (figura 24).

En respuesta a su reconocida relevancia, esta dimensión se centra en la introducción de medidas que preserven niveles razonables de validez y fiabilidad en el proceso de evaluación. Ampliamente documentados, los factores que pueden contaminar estos atributos en la implementación de los dispositivos ideados son muchos y variados. La falta de rigor o celo en el desarrollo del proceso de recogida de evidencias o las imprecisiones y desviaciones producidas durante su posterior análisis y valoración pueden dar lugar a conclusiones poco firmes y de calidad dudosa, tanto si pretenden emplearse con fines pedagógicos como acreditativos. Por ese motivo, el profesorado debe introducir medidas y mecanismos para contrarrestar o neutralizar de forma deliberada el impacto potencial de estos riesgos en la recogida de evidencias y su valoración (Biggs, 2005; Flemming, 2003; Leal, 2004). Máxime cuando la emergencia de las competencias como objeto evaluativo incrementa la complejidad del proceso evaluador, enfatizando la necesidad de efectuar juicios a partir de medidas indirectas y complementarias, recabadas en situaciones reales o verosímiles. De hecho, uno de los rasgos que caracteriza a las/os profesoras/es excelentes del estudio de Bain (2007) es que son capaces de buscar verdaderos indicadores de aprendizaje y evidencias que determinen con precisión que la/el estudiante ha desarrollado las competencias previstas.

Por limitaciones asociadas a la operatividad del modelo, esta dimensión no capta la verdadera complejidad de la validez y la fiabilidad; ésta supera ampliamente las clásicas comprobaciones estadísticas, especialmente si se pretende hacer de la evaluación una herramienta didáctica y no sólo un recurso para la acreditación de los resultados (De la Orden, 2011b; Mateo, 2000).

Supervisados ambos atributos en la recogida y análisis de evidencias, el profesorado debe configurar informes que reflejen la racionalidad y el rigor del proceso de evaluación, en especial, cuando ésta afecte a la calificación. Por influencia de los estándares de evaluación y de los referentes normativos e institucionales presentados en el capítulo anterior, el modelo ITC establece que dichos informes deben ser respetuosos con los derechos reconocidos al alumnado (anonimato e información, fundamentalmente) y ceñirse a las condiciones detalladas en los marcos reguladores.

b) Indicadores clave en su desarrollo.

El desarrollo de la dimensión quinta se concreta en tres indicadores de acuerdo a la descripción efectuada. En cada caso se establecen cinco descriptores de logro (**ANEXO 05**):

- *Articula deliberadamente medidas y mecanismos para supervisar la validez en el proceso de análisis e interpretación de evidencias.*

INADECUADO	SIN CONTROL DE VALIDEZ. No se valora expresamente la validez del sistema evaluativo puesto en práctica.
POCO ADECUADO	CONTROL INICIAL DE VALIDEZ. Toma ciertas medidas para garantizar que la evaluación no se desvíe de las competencias que debe valorar, pero sólo en el momento de su planificación (cuadros de cruce entre objetivos y exigencias evaluativas, etc.).
ADECUADO	VALIDEZ EN EL PROCESO. Toma medidas para favorecer que durante el análisis e interpretación de evidencias, la evaluación no se desvíe de las competencias que debe valorar (herramientas que centren la atención en los aspectos precisos a evaluar, corrección a ciegas, revisión de los cuadros de cruce iniciales, etc.).
MUY ADECUADO	CONTROL ALEATORIO. Realiza un control aleatorio para asegurar que las medidas adoptadas en defensa de la validez de la evaluación funcionan (segunda valoración de trabajos corregidos, revisión de valoraciones anteriores, etc.).
EXCELENTE	CONTROL ALEATORIO CON CONTRASTE. Controla de forma aleatoria la efectividad de las medidas adoptadas en defensa de la validez, discutiendo las conclusiones con colegas, superiores o asesores externos.

- *Articula deliberadamente medidas y mecanismos para supervisar la precisión en el proceso de análisis e interpretación de evidencias.*

INADECUADO	SIN CONTROL DE FIABILIDAD. No se valora expresamente la fiabilidad/precisión del sistema evaluativo puesto en práctica.
POCO ADECUADO	CONTROL INICIAL. Toma ciertas medidas para garantizar que la evaluación sea precisa en la valoración de las competencias a desarrollar, pero sólo en el momento de su planificación (triangulación o combinación de métodos, graduaciones de niveles de exigencia, etc.).
ADECUADO	FIABILIDAD EN EL PROCESO. Toma medidas para favorecer que, durante el análisis e interpretación de evidencias, la evaluación sea precisa determinando distintos niveles de competencia (medidas contra sesgos de orden, presentación o expectativas, herramientas de calificación que fijen estándares de medida uniformes, etc.).
MUY ADECUADO	CONTROL ALEATORIO. Realiza un control aleatorio para asegurar que las medidas adoptadas en defensa de la fiabilidad de la evaluación funcionan (segunda valoración de trabajos corregidos, agrupación de producciones por el nivel de logro evidenciado y revisión, etc.).
EXCELENTE	CONTROL ALEATORIO CON CONTRASTE. Controla de forma aleatoria la efectividad de las medidas adoptadas en defensa de la fiabilidad, discutiendo las conclusiones con colegas, superiores o asesores externos.

- *Presenta los resultados de forma clara, concisa y bien fundamentada, por medio de formatos que respetan la confidencialidad de las/os estudiantes.*

INADECUADO	SIN CONFIDENCIALIDAD. Comunica a las/os estudiantes sus resultados de forma escueta, sin fundamentarlos con aclaraciones, y en formatos que no siempre atienden plenamente a su confidencialidad (comunicación en público, referencia a identidades, etc.).
POCO ADECUADO	INSUFICIENTE DETALLE. Presenta los resultados en formatos que respetan la confidencialidad de las/os estudiantes, pero sin entrar en su fundamentación (sin escala de graduación u otras especificaciones que los aclaren).
ADECUADO	PRESENTACIÓN CLARA Y CONCISA. Presenta los resultados respetando la confidencialidad de las/os estudiantes y de forma clara y concisa (dando cuenta de la escala de valoración empleada y desglosando el dato si fuera oportuno).
MUY ADECUADO	POR CRITERIO. Presenta los resultados de forma confidencial, clara y concisa, fundamentando las valoraciones por medio de los criterios de evaluación aplicados.
EXCELENTE	POR CRITERIO Y COMPETENCIA. Presenta los resultados de forma confidencial, clara y concisa, fundamentando el nivel de logro evidenciado en cada competencia por medio de los criterios aplicados en su valoración.

c) Componentes integrantes.

Movilizados de forma armónica y conjunta según precise la complejidad de cada situación particular, algunos de los elementos más representativos que integran esta dimensión de la competencia evaluativa son los contemplados en la tabla 53:

Tabla 53
Principales componentes integrados en la interpretación y análisis de evidencias (dimensión 05)

INTERPRETACIÓN Y ANÁLISIS DE EVIDENCIAS		
BASE CONCEPTUAL Conocimientos, saberes, dominio de...	BASE PROCEDIMENTAL Capacidades en acción, desarrollo práctico	BASE ACTITUDINAL Actitudes, predisposiciones, principios, inquietudes
• Conceptos básicos de validez y fiabilidad. • Procedimientos de estimación de los niveles de validez y fiabilidad. • Amenazas frecuentes a la validez y la fiabilidad en la recogida de evidencias. • Amenazas frecuentes a la fiabilidad y validez en el proceso de análisis. • Fórmulas y mecanismos para la neutralización de sesgos y amenazas. • Normativa evaluativa.	• Adopción de medidas para reducir los sesgos asociados a las expectativas o el contraste (ocultar las identidades, desordenar los trabajos de modo que no ofrezcan indicios sobre la autoría, valorar la misma tarea en todo el grupo antes de emprender la revisión de otra, por ejemplo). • Cálculo o estimación de los niveles de fiabilidad y validez en instrumentos empleados. • Triangulación o contraste de resultados recabados por procedimientos diferentes.	• Supervisión, control, y autocontrol. • Rigor, profesionalidad, tecnicidad. • Sistematicidad. • Espíritu crítico, autocrítica. • Cautela, meticulosidad. • Creatividad e ingenio. • Orientación a la mejora. • Intercambio, juicio profesional. • Transparencia. • Respeto a los derechos del alumnado (información, confidencialidad).

-
• Formatos y procedimientos confidenciales para la comunicación de resultados.	• Realización de controles aleatorios para garantizar la adecuación y el rigor en el análisis de datos y evidencias. • Creación de herramientas de análisis o corrección (plantillas de valoración o rúbricas para centrar la atención en los aspectos que pretenden valorarse). • Supervisión de la validez de la evaluación en la recogida de evidencias y su análisis. • Supervisión de la fiabilidad de la evaluación en la recogida de evidencias y su análisis. • Discusión de las potencialidades y limitaciones de la evaluación desarrollada con colegas o superiores. • Identificación de los límites de la recogida y el análisis de evidencias para mostrar resultados de aprendizaje. • Elaboración de informes acordes a las necesidades y derechos de información del alumnado. • Presentación clara y confidencial de resultados de aprendizaje.
-

4.6 Regulación del aprendizaje y la enseñanza por medio de la evaluación.

Utilizar los resultados constatados en el proceso de evaluación para regular el aprendizaje y la enseñanza, identificando logros y áreas de mejora y estableciendo pautas de actuación oportunas para su superación.

a) Descripción de la competencia.

Concebida la evaluación como herramienta didáctica para la comprensión y la mejora, esta dimensión adopta una relevancia crucial en la actuación docente. Centra la atención en cómo emplea el profesorado la información generada en la evaluación para articular medidas que regulen⁷⁶ deliberadamente el aprendizaje

⁷⁶ Una de las acepciones que prevé la RAE (2014) para el verbo regular, es el ajuste del funcionamiento de un sistema a determinados fines, en nuestro caso, los aprendizajes previstos en las

discente en dirección a las metas formativas. Esta intervención puede tener lugar de forma directa sobre el trabajo del alumnado (feedback y proacción) o por medio de ajustes y adecuaciones centradas en la propia enseñanza.

La incorporación de esta sexta dimensión al modelo de competencia evaluativa responde al subproceso de feedback y proacción definido en el modelo ITC. Su puesta en práctica exige del profesorado tres actuaciones evaluativas clave para la estimulación consciente del aprendizaje, ordenadas en dos niveles progresivos de profundización: la comisión de retroinformación o feedback en primer lugar y, a partir de ahí, las medidas de proacción sobre el aprendizaje discente y/o sobre la enseñanza. Determinada desde el planteamiento teórico de la evaluación para el aprendizaje, la relevancia y pertinencia de esta dimensión es unánime en la literatura y en la investigación específica, y así lo ha demostrado el capítulo cuarto. Los antecedentes que se han aproximado a la delimitación de la competencia evaluativa del profesorado también coinciden en destacar esta dimensión, con independencia del nivel educativo o la titulación en la que vaya a ejercerse (tabla 46). No sucede así en las escalas administradas al alumnado para la valoración de la calidad docente y en los instrumentos empleados para indagar sobre la evaluación del aprendizaje universitario; en ellas, su reflejo es reducido y se limita a cuestiones prácticas como la frecuencia, más que al efecto que provoca sobre el aprendizaje.

El primer nivel para ejercer la regulación del aprendizaje es el feedback, que puede adoptar formatos y modalidades diferentes según lo expuesto en el capítulo anterior. El modelo ITC integra ocho condiciones para guiar de forma efectiva esta retroinformación (criterio 5.2) y la delimitación de esta sexta dimensión de la competencia evaluativa incorpora en su desarrollo la mayor parte de las mismas. No obstante, por razones prácticas asociadas a la manejabilidad de la propuesta, dos de ellas no han sido explícitamente contempladas entre los descriptores de logro:

- El feedback puede proceder de fuentes diversas (Carless et al., 2011; Dochy et al., 1999; Falchikov, 1995; Hounsell, 2007; Liu & Carless, 2006; Nicol et al., 2014), sin embargo esta sexta dimensión hace referencia a la información y las pautas de mejora generadas intencionalmente por el profesorado⁷⁷.

competencias. Así lo plantean en sus trabajos Mateo (2000), Sanmartí (2007) y Monereo y Castelló (2009a), para quienes regular, es adecuar e incidir deliberadamente, no poner normas (otra de las acepciones del término).

⁷⁷ Esta acotación ya ha sido justificada en el capítulo cuarto al clarificar la visión del feedback. Así pues, la generación de feedback sobre la propia actuación o los procesos de feedback entre iguales quedan vinculados a la próxima dimensión, relativa a la participación discente en la evaluación.

- Unido a ello, el carácter dialógico representa otro rasgo asociado a la calidad del feedback (Carless, 2013b; Nicol, 2010; Sadler, 2010) que tampoco se ha trasladado expresamente a la rúbrica que define la competencia evaluativa. Considerando que los objetivos de la investigación son generales (delimitar la competencia evaluativa, no el feedback en particular), se ha optado por contemplar otros rasgos más amplios de la retroalimentación.

La proacción tiene lugar a partir del feedback generado y está relacionada con su carácter prospectivo, orientado a la mejora del aprendizaje. En este sentido, el profesorado debería ser capaz de revertir la comprensión que proporciona la evaluación en la mejora del aprendizaje, recurriendo a dos vías compatibles: pautas y apoyos directos para ayudar al alumnado en su progreso o en ajustes que afecten a su enseñanza, para procurar una mejor contribución al aprendizaje.

b) Indicadores clave en su desarrollo.

La sexta dimensión incorpora tres indicadores interrelacionados y cinco descriptores de logro para cada uno de ellos (**ANEXO 05**):

- *Ofrece a las/os estudiantes feedback a partir de las actividades evaluativas desarrolladas.*

INADECUADO	COMUNICACIÓN RESULTADOS. La información que devuelve a las/os estudiantes sobre su aprendizaje y su trabajo es básicamente cuantitativa (por lo general la calificación adjudicada).
POCO ADECUADO	COMENTARIOS ESPORÁDICOS. Devuelve a las/os estudiantes algunos comentarios sobre su aprendizaje y su trabajo, pero con cierta demora y no siempre (de vez en cuando o sólo si la solicitan).
ADECUADO	FEEDBACK FRECUENTE, PUNTUAL y SUFICIENTE. Proporciona a las/os estudiantes información frecuente, suficiente y puntual sobre su aprendizaje y su trabajo, para que comprendan la valoración recibida y los resultados esperados.
MUY ADECUADO	...Y PERSONALIZADO. Proporciona a las/os estudiantes información personalizada, frecuente, suficiente y puntual sobre su aprendizaje y su trabajo, para que cada una/o de ellas/os comprenda el punto en el que se encuentra y los resultados esperados.
EXCELENTE	...Y ESTIMULANTE. Proporciona a las/os estudiantes información personalizada, frecuente, suficiente y puntual sobre su aprendizaje y su trabajo que, centrada en los aspectos positivos, constituye un estímulo para seguir avanzando en línea con los resultados esperados.

- *Especifica pautas claras de proacción para orientar a las/os estudiantes en el desarrollo de las competencias previstas.*

INADECUADO	COMUNICACIÓN RESULTADOS. La información que devuelve a las/os estudiantes sobre su aprendizaje y su trabajo es básicamente cuantitativa, rara vez va acompañada de sugerencias de mejora o pautas de actuación.
POCO ADECUADO	SUGERENCIA PUNTUAL DE PAUTAS GENERALES. Puntualmente (de vez en cuando o si se lo solicitan), propone a las/os estudiantes sugerencias o pautas generales de actuación para mejorar los resultados de aprendizaje observados.
ADECUADO	PROACCIÓN MEJORA. Especifica pautas claras de actuación y mejora para orientar al grupo de estudiantes en el desarrollo de las competencias en función de los resultados de aprendizaje observados.
MUY ADECUADO	PROACCIÓN PERSONALIZADA. Especifica pautas claras de actuación y mejora para orientar a las/os estudiantes de forma personalizada en el desarrollo de las competencias en función de sus resultados de aprendizaje.
EXCELENTE	PROACCIÓN y SEGUIMIENTO. Realiza seguimiento de la aplicación de las pautas de mejora proporcionadas a cada estudiante en función de sus resultados para el logro de las competencias.

- *Introduce ajustes para mejorar la enseñanza en función de los resultados de aprendizaje de las/os estudiantes.*

INADECUADO	SIN REVISIÓN. Los resultados de aprendizaje de las/os estudiantes no se utilizan para introducir ajustes o replantear la docencia.
POCO ADECUADO	REVISIÓN NO SISTEMÁTICA. Los resultados de aprendizaje de las/os estudiantes plantean dudas e inquietudes que tienen que ver con la forma de enseñar.
ADECUADO	REVISIÓN y AJUSTE EN FUNCIÓN DE RESULTADOS. Se introducen ajustes y modificaciones en la forma de enseñar en función de los resultados de aprendizaje de las/os estudiantes.
MUY ADECUADO	... Y SUGERENCIAS. Se introducen ajustes y modificaciones en la enseñanza en función de los resultados de aprendizaje de las/os estudiantes y las recomendaciones y sugerencias que directamente se les solicitan al respecto.
EXCELENTE	REVISIÓN, AJUSTE y SEGUIMIENTO DE LA DOCENCIA. Documenta los resultados de aprendizaje discente y las sugerencias de mejora registradas año tras año, para reflexionar y hacer seguimiento de los puntos fuertes y áreas de mejora vinculados a la docencia y su evolución.

c) *Componentes integrantes.*

Movilizados de forma armónica y conjunta según precise la complejidad de cada situación particular, algunos de los elementos más representativos que integran esta dimensión de la competencia evaluativa docente son los presentados en la tabla 54:

Tabla 54

Principales componentes integrados en la regulación del aprendizaje y la enseñanza (dimensión 06)

REGULACIÓN DEL APRENDIZAJE Y LA ENSEÑANZA		
BASE CONCEPTUAL Conocimientos, saberes, dominio de...	BASE PROCEDIMENTAL Capacidades en acción, desarrollo práctico	BASE ACTITUDINAL Actitudes, predisposiciones, principios, inquietudes
• Técnicas y modalidades de feedback. • Condiciones bajo las cuales el feedback contribuye al aprendizaje. • Dificultades frecuentes del alumnado ante el feedback (decodificación y diferente percepción). • Impacto o repercusión emocional del feedback. • Los resultados de aprendizaje como evidencia de la efectividad docente. • Repertorio de recursos didácticos y metodológicos (actividades, dinámicas, recursos, etc. que puedan ayudar al alumnado a reducir la distancia entre la situación actual y las metas formativas).	• Combinación viable y efectiva de diversas modalidades de feedback. • Comunicación clara, adecuada y descriptiva de la situación actual y de las metas de aprendizaje. • Establecimiento de conexiones claras entre los logros evidenciados y los pretendidos. • Identificación de errores y lagunas clave que interfieren en el aprendizaje. • Ajuste de las representaciones discentes del aprendizaje a realizar. • Proporción puntual de información sobre el avance y pautas de regulación pertinente. • Estimulación positiva del trabajo del alumnado. • Revisión constructiva de los resultados de aprendizaje en términos de puntos fuertes y áreas de mejora. • Formulación de propuestas claras de actuación a partir de los resultados evidenciados y su análisis. • Propuesta personalizada de alternativas de mejora y actuación. • Revisión y ajuste de la enseñanza a partir de los resultados constatados. • Investigación sobre la propia efectividad docente. • Formulación y seguimiento de retos de desarrollo profesional.	• Orientación al aprendizaje. • Orientación a la mejora. • Sensibilidad, empatía. • Diálogo, apertura. • Persuasión. • Creatividad e ingenio. • Autocrítica. • Desarrollo profesional.

4.7 Participación de las/os estudiantes en la evaluación.

Utilizar la participación activa de las/os estudiantes en el proceso evaluativo para contribuir a la autorregulación progresiva de su aprendizaje hacia las metas formativas definidas.

a) Descripción de la dimensión.

La séptima dimensión refiere a la capacidad del profesorado para promover deliberadamente experiencias evaluativas participativas que favorezcan en el alumnado la consecución de los objetivos de aprendizaje previstos. Partiendo del protagonismo natural de las/os estudiantes en la construcción de su propio aprendizaje, esta dimensión plantea su intervención en las cuatro fases de la evaluación: diseño, implementación, establecimiento de conclusiones y/o en la revisión⁷⁸. Esta participación debe ser planteada conscientemente por el profesorado, como una experiencia más de aprendizaje para la consecución de los fines previstos. Según los propósitos concretos de cada situación, puede combinar los distintos formatos de evaluación participativa definidos (**ANEXO 03**), con efecto en la calificación (autocalificación, co-calificación, calificación dialogada) o independiente de ella (autoevaluación, coevaluación, evaluación compartida). Resulta especialmente pertinente entre el profesorado vinculado a la formación básica de maestras/os, ya que lleva a situar al alumnado en un rol que deberá ejercer en su futuro profesional (I. Álvarez, 2008a; C. Brew et al., 2009; V. M. López, Castejón, et al., 2012; McGarr & Clifford, 2013).

La implicación discente en la evaluación del aprendizaje es uno de los criterios de calidad definidos en el modelo ITC. Según lo detallado para su justificación, la participación del alumnado representa uno de los principales ejes de actuación dentro del enfoque teórico de la evaluación para el aprendizaje y una de las repercusiones evaluativas asociadas al EEES. Su traslado a esta séptima dimensión coincide con una de las grandes funciones evaluativas delimitadas por Quesada et al. (2013) para el profesorado universitario y con los indicadores propuestos por Toledo (2005, 2006), Stiggins (1998) y J. Chappuis et al. (2012) para la actuación docente. Sin embargo, como en los referentes normativos y en las propuestas más institucionales para la calidad de la evaluación, su representación

⁷⁸ Ese es el motivo de que haya pasado a conformar una dimensión diferenciada, a pesar de que el modelo ITC incluya la participación entre las decisiones asociadas a la planificación estratégica de dispositivos evaluativos.

es escasa en las escalas destinadas a valorar la calidad de la docencia (figura 24). Las alusiones a las oportunidades brindadas al alumnado para intervenir en la evaluación son minoritarias, lo que sugiere un menor reconocimiento de su importancia. Salvo en los estudios que han procurado indagar expresamente en la participación discente en la evaluación (Arribas et al., 2010; V. M. López & Palacios, 2012; G. Rodríguez et al., 2012), así se evidencia también entre las herramientas aplicadas en la investigación sobre la praxis evaluativa en la universidad.

b) Indicadores clave en su desarrollo.

El desarrollo de la penúltima dimensión se concreta en un único indicador, asociado a cinco descriptores de realización diferentes (**ANEXO 05**):

- *Desarrolla actividades que provocan la participación activa de las/os estudiantes en la evaluación.*

INADECUADO	SIN PARTICIPACIÓN DISCENTE. En el sistema de evaluación desarrollado la participación de las/os estudiantes se limita a la de agentes evaluados principalmente.
POCO ADECUADO	PARTICIPACIÓN PUNTUAL. De forma puntual se plantean a las/os estudiantes actividades de evaluación participativa (autoevaluación y/o evaluación entre compañeras/os) que permiten reducir la carga de corrección y los tiempos de devolución de respuesta.
ADECUADO	EVALUACIÓN PARTICIPATIVA. Desarrolla actividades de evaluación participativa, donde las/os estudiantes profundizan en la comprensión de los criterios de evaluación, al aplicarlos a la valoración de su propio trabajo y/o el de sus compañeras/os.
MUY ADECUADO	PARTICIPATIVA: DISEÑO y APLICACIÓN. Implica a las/os estudiantes en la definición de las condiciones y requisitos de la evaluación, y en la posterior aplicación de los mismos a la valoración de su propio trabajo y/o el de sus compañeras/os.
EXCELENTE	...Y REVISIÓN. Implica a las/os estudiantes en el diseño y desarrollo del sistema evaluativo, así como en la revisión de las principales ventajas y limitaciones del mismo.

Esta séptima dimensión concentra un único componente, igual que la segunda. A pesar de los riesgos metodológicos que entraña, se ha optado por definir un único indicador para simplificar el modelo y evitar la fatiga en el manejo de los instrumentos generados a partir del mismo. Relacionada con la planificación estratégica y con la regulación del aprendizaje, esta decisión permite explorar las relaciones que aprecian docentes y discentes entre la evaluación participativa y otras dimensiones o componentes del proceso.

c) Componentes integrantes.

Movilizados de forma armónica y conjunta según precise la complejidad de cada situación particular, algunos de los elementos más representativos que integran esta dimensión de la competencia evaluativa docente son los ordenados en la tabla 55:

Tabla 55

Principales componentes integrados en la interpretación y análisis de evidencias (dimensión 07)

PARTICIPACIÓN DE LAS/OS ESTUDIANTES EN LA EVALUACIÓN		
BASE CONCEPTUAL Conocimientos, saberes, dominio de...	BASE PROCEDIMENTAL Capacidades en acción, desarrollo práctico	BASE ACTITUDINAL Actitudes, predisposiciones, principios, inquietudes
• Modalidades de evaluación participativa. • Posibilidades de intervención discente en el proceso de evaluación. • Usos y beneficios potenciales de la evaluación participativa. • Principales amenazas y dificultades de la evaluación participativa. • Resistencias habituales del alumnado ante la participación evaluativa. • Repertorio de procedimientos, técnicas y estrategias de evaluación participativa.	• Diseño de experiencias evaluativas participativas. • Desarrollo de experiencias evaluativas para la consecución de los fines formativos previstos. • Implicación activa del alumnado en las distintas fases de la evaluación. • Exposición clara de los términos y los objetivos de la participación. • Detección de los riesgos y las malas prácticas durante el desarrollo de experiencias evaluativas participativas. • Previsión de riesgos y dificultades asociadas a la evaluación participativa. • Estimulación consciente de procesos de generación de feedback sobre la propia actuación o entre iguales. • Estimulación deliberada de la competencia evaluativa de las/os futuras/os maestras/os. • Creación de situaciones específicamente ideadas para situar al alumnado en el rol evaluador. • Clarificación de los criterios de evaluación y expectativas de aprendizaje para favorecer su ejercicio racional entre el alumnado. • Orientación formativa de la experiencia participativa en el trabajo y aprendizaje individual futuro.	• Confianza en la capacidad del alumnado para implicarse de forma responsable en el proceso de evaluación. • Cautela. • Diálogo, comunicación, apertura. • Negociación y consenso. • Orientación a la mejora y el aprendizaje, desarrollo profesional.

4.8 Revisión del dispositivo evaluativo.

Revisar el sistema de evaluación para identificar fortalezas y debilidades que puedan orientar su mejora desde la perspectiva ética, técnica y pedagógica.

a) Descripción de la dimensión.

La metaevaluación de los dispositivos evaluativos representa un subproceso asociado a la calidad de la evaluación según la definición establecida en el modelo ITC. En consecuencia, la competencia evaluativa exige que el profesorado someta a revisión y valore de forma racional y sistemática la calidad de los sistemas evaluativos diseñados y puestos en práctica para la estimulación y acreditación del aprendizaje. Este análisis representa la cuarta fase del proceso de evaluación, y puede desarrollarse durante la puesta en práctica de los dispositivos articulados y/o una vez finalizada. Independientemente del momento en el que se desarrolle, debe reparar en las dimensiones técnica, ética y pedagógica de la evaluación, de acuerdo con los requisitos establecidos en el anterior capítulo. En este proceso, se considera deseable que el profesorado recurra a fuentes de información que complementen su punto de vista, para mayor riqueza y rigor en la revisión. Por su posición con respecto al objeto de análisis, resulta de particular interés recabar el testimonio de estudiantes y de colegas, aunque se puede recurrir también a colaboraciones procedentes de auditoría externa.

Reconocida como un medio eficaz para valorar y contribuir a la calidad de la evaluación (De Miguel, 2010; Moreno & Santos, 2004; Pattel, 2002; Stake, 2006; Stufflebeam, 2001b, 2011/1974), la capacidad de metaevaluación se encuentra ampliamente representada en los referentes consultados para la definición de la competencia evaluativa docente. Por una parte, coincide con la revisión, una de las funciones evaluativas clave delimitadas por Quesada et al. (2013). Por la otra, la idea de metaevaluación subyace a todo intento de definición de normas o estándares para dirigir la praxis evaluativa del profesorado (American Association of School Administrators, National Association of Elementary School Principals, National Association of Secondary School Principals, & National Council on Measurement in Education, 1997; American Federation of Teachers, National Council on Measurement in Education, & la National Education Association, 1990; Monereo & Castelló, 2009a). Además, en el contexto específico de la formación básica de maestras/os adopta una significación especial, como sugiere Santos (2003):

Las instituciones de formación de profesores tienen una exigencia más poderosa que cualquier otra para poner en marcha estos procesos de reflexión, ya que son instituciones, que plantean en su currículum la reflexión teórica sobre estas cuestiones. Resulta contradictorio explicar a los alumnos y alumnas la importancia que tiene la evaluación y cerrarse a su realización por temor, pereza o dogmatismo. (p. 83).

A pesar de que su relevancia resulte evidente desde la perspectiva teórica, la metaevaluación no figura entre las cuestiones contempladas habitualmente en las herramientas aplicadas a la investigación evaluativa, ni en las escalas para la evaluación de la calidad docente. Aun así, sostenemos que el profesorado debe mostrar solvencia en la revisión de los dispositivos evaluativos articulados. El resultado de esta metaevaluación debe ser la identificación de las posibilidades y limitaciones de la evaluación desarrollada con respecto a la promoción y acreditación del aprendizaje discente, de modo que puedan articularse propuestas y ajustes para su aproximación a la visión definida en el modelo ITC.

b) Indicadores clave en su desarrollo.

La última dimensión de la competencia evaluativa docente se concreta en tres indicadores interrelacionados. En cada uno de ellos se presentan descriptores para cinco niveles de realización ascendentes (**ANEXO 05**):

- *Revisa el cumplimiento de los principios de utilidad y consecuencialidad en el sistema evaluativo experimentado.*

INADECUADO	NO REvisa. No lleva a cabo una revisión específica de la utilidad que tiene el sistema evaluativo para condicionar de forma positiva el aprendizaje de las/os estudiantes.
POCO ADECUADO	REVISIÓN NO SISTEMÁTICA. Reflexiona sobre la utilidad que tiene el sistema evaluativo aplicado para condicionar de forma positiva el aprendizaje de las/os estudiantes, pero no de forma sistemática.
ADECUADO	REVISIÓN SISTEMÁTICA. Revisa de forma sistemática la utilidad que tiene el sistema evaluativo para condicionar de forma positiva el aprendizaje de las/os estudiantes, introduciendo ajustes y correcciones que contribuyan a su mejora.
MUY ADECUADO	REVISIÓN CON ESTUDIANTES. Revisa de forma sistemática y en colaboración con las/os estudiantes la utilidad que tiene el sistema evaluativo para condicionar su aprendizaje de forma positiva, introduciendo ajustes y correcciones que contribuyan a su mejora.

EXCELENTE	REVISIÓN CON ESTUDIANTES Y OTRO AGENTE. Revisa de forma sistemática y en colaboración con las/os estudiantes y al menos otro agente externo (asesor/a, colega, etc.) la utilidad que tiene el sistema evaluativo aplicado para condicionar de forma positiva el aprendizaje, introduciendo ajustes y correcciones para su mejora.

- *Revisa el cumplimiento de los principios de viabilidad, sistematicidad, validez y fiabilidad en el sistema evaluativo.*

INADECUADO	NO REvisa. No lleva a cabo una revisión específica de la viabilidad y sistematicidad del sistema evaluativo para producir valoraciones precisas de los aprendizajes previstos.
POCO ADECUADO	REVISIÓN NO SISTEMÁTICA. Reflexiona sobre la viabilidad y sistematicidad del sistema evaluativo para producir valoraciones precisas de los aprendizajes previstos, pero no de forma sistemática.
ADECUADO	REVISIÓN SISTEMÁTICA. Revisa de forma sistemática la viabilidad y sistematicidad del sistema evaluativo para producir valoraciones precisas, introduciendo ajustes y correcciones que contribuyan a su mejora.
MUY ADECUADO	REVISIÓN CON ESTUDIANTES. Revisa de forma sistemática y en colaboración con las/os estudiantes la viabilidad y sistematicidad del sistema evaluativo para producir valoraciones precisas, introduciendo ajustes y correcciones que contribuyan a su mejora.
EXCELENTE	REVISIÓN CON ESTUDIANTES Y OTRO AGENTE. Revisa de forma sistemática y en colaboración con las/os estudiantes y al menos otro agente externo (asesor/a, colega, etc.) la viabilidad, sistematicidad, validez y fiabilidad del sistema evaluativo, introduciendo ajustes y correcciones para su mejora.

- *Revisa el cumplimiento de los principios éticos de justicia, privacidad, transparencia y beneficencia en el sistema evaluativo.*

INADECUADO	NO REvisa. No lleva a cabo una revisión específica de la justicia, transparencia, privacidad y beneficencia del sistema evaluativo experimentado.
POCO ADECUADO	REVISIÓN NO SISTEMÁTICA. Reflexiona sobre la justicia, transparencia, privacidad y beneficencia del sistema evaluativo experimentado, pero no de forma sistemática.
ADECUADO	REVISIÓN SISTEMÁTICA. Revisa de forma sistemática la justicia, transparencia, privacidad y beneficencia del sistema evaluativo, introduciendo ajustes y correcciones que contribuyan a su mejora.
MUY ADECUADO	REVISIÓN CON ESTUDIANTES. Revisa de forma sistemática y en colaboración con las/os estudiantes la justicia, transparencia, privacidad y beneficencia del sistema evaluativo, introduciendo ajustes y correcciones que contribuyan a su mejora.
EXCELENTE	REVISIÓN CON ESTUDIANTES Y OTRO AGENTE. Revisa de forma sistemática y en colaboración con las/os estudiantes y al menos otro agente externo (asesor, colega, etc.) la justicia, transparencia, privacidad y beneficencia del sistema evaluativo, introduciendo ajustes y correcciones para su mejora.

c) Componentes integrantes.

Movilizados de forma armónica y conjunta según precise la complejidad de cada situación particular, algunos de los elementos más representativos que integran esta dimensión de la competencia evaluativa docente son los mostrados en la tabla 56:

Tabla 56

Principales componentes integrados en la revisión del dispositivo evaluativo (dimensión 08)

REVISIÓN DEL DISPOSITIVO EVALUATIVO		
BASE CONCEPTUAL Conocimientos, saberes, dominio de...	BASE PROCEDIMENTAL Capacidades en acción, desarrollo práctico	BASE ACTITUDINAL Actitudes, predisposiciones, principios, inquietudes
• Métodos y modalidades de metaevaluación. • Principios pedagógicos de referencia para el desarrollo de prácticas evaluativas. • Principios técnicos de referencia para el desarrollo de prácticas evaluativas. • Principios éticos de referencia para el desarrollo de prácticas evaluativas. • Limitaciones de los sistemas de evaluación y beneficios de la revisión.	• Recopilación de evidencias sobre la adecuación ética, técnica y pedagógica del sistema evaluativo. • Registro de las incidencias que tienen lugar durante la puesta en práctica del dispositivo evaluativo. • Implicación del alumnado en la revisión de la evaluación. • Colaboración con agentes externos en la revisión de la evaluación. • Evaluación del grado de adecuación ética, técnica y pedagógica de la evaluación. • Reconocimiento de prácticas evaluativas nocivas, poco éticas y/o fiables. • Análisis de alternativas de actuación. • Selección de las evidencias necesarias para la metaevaluación y de las fuentes e instrumentos para su recopilación efectiva. • Identificación de los límites y las posibilidades del dispositivo evaluativo empleado. • Formulación de ajustes y propuestas de mejora a partir de los resultados.	• Orientación a la mejora. • Espíritu crítico, autocrítica. • Revisión, reflexión y ajuste. • Profesionalidad, formación permanente. • Diálogo y colaboración. • Sistematicidad, rigor. • Sentido ético.

5. El capítulo a grandes rasgos

El quinto capítulo del marco teórico ha presentado el modelo de competencia evaluativa propuesto para el profesorado universitario vinculado a la formación básica de maestras/os de Educación Infantil y/o Primaria. Ideada con fines formativos y/o evaluativos, dicha propuesta estructura la competencia evaluativa docente en ocho dimensiones y elabora por cada una de ellas un sistema de indicadores con cinco descriptores de logro para cada caso. El resultado es una rúbrica que puede aplicarse como hoja de ruta para el diseño y desarrollo de dispositivos evaluativos de calidad y/o como recurso para la metaevaluación.

El establecimiento de las dimensiones, indicadores y descriptores del modelo se ha realizado en un proceso con siete grandes etapas de trabajo. En su desarrollo se han integrado e incorporado de forma reflexiva las aportaciones derivadas de la consulta a un grupo de expertas/os y el análisis de cuatro referentes relevantes:

- El *modelo ITC de evaluación* presentado en el capítulo cuarto, que define las prácticas evaluativas de calidad en el marco de la investigación en curso.
- *Otras propuestas que se han detenido en la delimitación de la competencia evaluativa del profesorado.* Esta indagación se ha efectuado sobre tres tipos de antecedentes, ordenados según su proximidad a los propósitos de la investigación en curso: la delimitación de la competencia evaluativa docente en el contexto de la formación básica de maestras/os de Educación Infantil y Primaria, en la enseñanza universitaria, o en la docencia en general. Esta revisión de los intentos previos de sistematización se ha llevado a cabo en la fase previa de definición del modelo, y en la fase final de la investigación con fines confirmatorios.
- La consideración de 30 *herramientas de probada validez empleadas en el análisis de las prácticas evaluativas desarrolladas en la enseñanza universitaria.*
- Las conclusiones de un trabajo de campo efectuado sobre los *tópicos evaluativos incluidos en las escalas de valoración de la calidad docente administradas en los grados vinculados a la formación básica de maestras/os.* El trabajo se ha llevado a cabo sobre un total de 63 escalas, en representación del 93% de las universidades y del 77% de los centros de formación superior que imparten los Grados de Educación Infantil y Educación Primaria en todo el estado.

La consulta de referentes ha dado lugar a la delimitación de un modelo que sistematiza en 19 indicadores la competencia evaluativa del profesorado vinculado a la formación básica de maestras/os. El resultado no reúne los requisitos de rigor y consenso para ser considerado un sistema ordenado de estándares (Baeza et al., 2006; JCSEE, 2007). Sin embargo, con los matices y reservas que exige su provisionalidad, sus potencialidades y contribuciones coinciden, a grandes rasgos, con los señalados en la literatura para los estándares profesionales. De acuerdo con Ingvarson y Kleinhenz (2006) los estándares profesionales establecen “los valores centrales de la educación que guían a los profesores eficaces e identifican las características y tareas exclusivas de la enseñanza como profesión. [...], dejan claro qué necesitan saber y hacer los profesores para mejorar a largo plazo” (p. 269). De modo similar, la definición de un modelo que delimita la competencia evaluativa docente favorece:

- *El desarrollo de sistemas evaluativos de calidad*, según lo detallado en el modelo ITC de evaluación.
- *La revisión y la mejora de la praxis evaluativa*, aportando un criterio comparativo desde el cual poder detectar los principales puntos fuertes y áreas de mejora.
- *La autorregulación y búsqueda de la excelencia profesional y la formación permanente del profesorado*, ofreciendo un marco de referencia y una herramienta para revisar y encauzar la praxis evaluativa y estimular con ello el aprendizaje y desarrollo profesional.

Frente a otras aproximaciones que han tratado de acercarse a la praxis evaluativa del profesorado, el modelo de competencia resultante presenta los siguientes beneficios o ventajas potenciales:

- Tanto si se emplea como recurso formativo o evaluativo, se trata de una propuesta flexible, abierta a posibilidades diversas de aplicación según se requiera en cada caso. Partiendo del principio de voluntariedad y aplicación pedagógica (no acreditativa), las experiencias de formación y evaluación pueden desarrollarse a distintos niveles (interno o externo, individual o grupal, etc.) y con implicación de diversos agentes (profesorado, estudiantes, colegas y/o expertos).
- El modelo centra la atención en el ámbito de actuación docente y recorre todas aquellas decisiones y actuaciones evaluativas que competen de forma directa al profesorado. Sin negar la relevancia e incidencia de la perspectiva institucional (reglamentos y normativas evaluativas, rendición de cuentas, apoyos y servicios, etc.), las dimensiones, indicadores y descriptores propuestos aterrizan en el nivel

más directo e inmediato del proceso de evaluación, aquel que tiene lugar en la interacción didáctica con el alumnado, para estimular y/o acreditar su aprendizaje.

El desarrollo del marco empírico de la investigación se aproxima a la competencia evaluativa del profesorado tal y como ha quedado definida en el presente capítulo. En concreto, propone un diagnóstico individual y externamente dirigido de la misma, contando con la participación de estudiantes y profesorado.

CAPÍTULO 6.

PLANTEAMIENTO TEÓRICO DE LA COMPETENCIA EVALUATIVA DEL PROFESORADO: CONCLUSIONES DEL MARCO TEÓRICO

En consonancia con lo establecido en el planteamiento general de la investigación, una de las motivaciones centrales a las que responde el presente trabajo es la identificación de los rasgos que deben caracterizar la competencia evaluativa docente en los grados vinculados a la formación básica de maestras/os de Educación Infantil y/o Primaria. Su respuesta ha sido articulada por medio del marco teórico, en el que han sido atendidos de forma sucesiva los cuatro propósitos definidos para su consecución.

Antes de pasar a abordar la competencia evaluativa del profesorado desde la perspectiva empírica, este último capítulo rescata las conclusiones más relevantes extraídas de la aproximación teórica. Plantea para ello una discusión preliminar de los resultados asociados a cada uno de los cuatro objetivos que la han guiado, antes de su integración final en las conclusiones generales de la investigación.

1. OBJETIVO A. Formular una propuesta revisada de las principales condiciones y características que debe reunir la evaluación del aprendizaje en el contexto de la formación básica de maestras/os

La importancia y pertinencia de la evaluación del aprendizaje universitario como objeto de estudio representa un acuerdo unánime sobre el que no cabe duda. Afecta a una amplia gama de audiencias interesadas y constituye un núcleo en el que convergen inquietudes técnicas, éticas, pedagógicas o legales. Por todo ello, destaca como un foco de

atención recurrente, con una presencia cada vez mayor en los foros y equipos de investigación (I. Álvarez, 2008b; Ibarra & Rodríguez, 2010).

El interés natural que suscita el tema se traduce en la publicación de numerosas referencias. Desde diferentes perspectivas y con desigual nivel de sistematización, existen múltiples propuestas que aportan orientaciones para el diseño, la puesta en práctica y/o la revisión de la evaluación del aprendizaje. Las pautas y principios de actuación resultantes muestran un grado de abstracción y desarrollo muy variable, desde la formulación de orientaciones generales, a la delimitación de estándares. También reflejan visiones particulares del fenómeno evaluativo, dependiendo de la perspectiva desde la que hayan sido elaboradas y de la audiencia a la que representen prioritariamente.

A pesar de este amplio desarrollo (o incluso unido al mismo), no existe una visión compartida sobre la calidad de la evaluación del aprendizaje en la universidad (Buendía & López, 1999; Cano et al., 2010; Porto, 2008), lo que dificulta severamente la coordinación y la unidad de criterio en este terreno (Briggs et al., 2003; Gairín, 2009; B. Salinas & Cotillas, 2007). En consecuencia, el primer objetivo de la investigación ha sido formular una propuesta fundamentada de las principales características y condiciones que debe reunir la evaluación del aprendizaje en el contexto específico de la formación básica de maestras/os.

La respuesta a este primer reto se ha materializado en el modelo de evaluación ITC (Integrado Teórico Complejo), presentado en el capítulo cuarto. Posicionado en la perspectiva teórica de la evaluación para el aprendizaje, delimita un sistema de siete subprocesos, 19 criterios y 68 indicadores para definir la calidad evaluativa. En el marco de la presente investigación el modelo ITC representa la teoría evaluativa de referencia: una visión parcial del fenómeno evaluativo, construida sobre referentes relevantes que le dotan de respaldo y validez.

La principal contribución de la propuesta ITC es que estructura una síntesis operativa de diversas orientaciones para el ejercicio de la evaluación. No obstante, no todos los criterios e indicadores que integra gozan del mismo respaldo entre las fuentes consideradas en su elaboración. Explorados en el capítulo cuarto, los referentes adoptados en su construcción se aproximan a la evaluación desde la perspectiva teórica, institucional o normativa, fundamentalmente. En consecuencia, coinciden en señalar la relevancia de determinados rasgos evaluativos y toman distancia en otros tantos.

Con independencia de la fuente consultada, existen dos rasgos o condiciones evaluativas que suscitan un acuerdo unánime:

- *Explicitación y difusión de criterios y condiciones.* Frente a las prácticas evaluativas opacas e implícitas generalizadas en la tradición evaluativa (Darby, 2007), la

necesidad de transparencia y claridad se señala sin excepción desde las diversas perspectivas. Mientras que los referentes institucionales y normativos insisten en la satisfacción del derecho de información de las/os estudiantes, las aportaciones teóricas ahondan en los beneficios formativos que reporta la interiorización de las expectativas de aprendizaje y la comprensión de las condiciones para su valoración.

- *Continuidad de la evaluación y valor formativo.* Frente a la tradición evaluativa final y sumativa descrita en el capítulo tercero, los referentes consultados contemplan la necesidad de hacer uso de la función pedagógica de la evaluación a lo largo de todo el proceso formativo. Por su orientación eminentemente técnica, las propuestas estandarizadas representan los únicos referentes que no recogen expresamente esta condición, aunque sí cuestiones próximas a ella (dentro de los estándares de utilidad, información sobre usos y principios de referencia, seguimiento, etc.).

En cuanto a las discrepancias, los referentes manejados muestran dos grandes posiciones u orientaciones de partida con respecto a la calidad de la evaluación, que les llevan a reparar en cuestiones diferentes:

- *Orientación pedagógica de la evaluación.* Esta primera visión se hace especialmente presente en los decálogos y manifiestos que tratan de guiar la praxis evaluativa. Claramente inspirados en los principios de la perspectiva teórica de la evaluación para el aprendizaje, ofrecen una visión didáctica de la misma e insisten en su potencialidad pedagógica. Desde esta orientación existen dos aspectos concretos que encuentran un respaldo unánime por su contribución a la construcción del aprendizaje: el proceso de feedback o retroalimentación y la participación del alumnado en el proceso evaluativo.
- *Orientación técnica-normativa de la evaluación.* La segunda orientación se hace manifiesta en los referentes normativos, en las propuestas estandarizadas, así como en las directrices institucionales para orientar la praxis evaluativa. En ella la calidad de la evaluación se define en referencia a cuestiones técnicas y al cumplimiento de las disposiciones normativas, más que desde la perspectiva pedagógica. Confirmando las apreciaciones de Verdejo (2008), Ibarra y Rodríguez (2010) y Padilla et al. (2010), los instrumentos y procedimientos evaluativos aplicados reciben una atención prioritaria desde esta orientación, destacando su variedad e idoneidad técnica para la acreditación válida y fiable del aprendizaje. También se subraya el cumplimiento de los derechos y deberes ligados a la evaluación, adoptando como principal referencia el punto de vista discente. Junto a cuestiones como la

protección de la identidad o la justicia (entendida como igualdad y objetividad), la inquietud por la transparencia enlaza con el derecho a la información en este caso.

La lectura vertical de las directrices institucionales, los referentes normativos y las propuestas estandarizadas sugieren que, en cierta medida, persiste la clásica confusión entre evaluación y calificación (Andreu-Andrés & Labrador-Piquer, 2012; V. M. López, 2004). Lejos del planteamiento de Villa (2004), Kaneko (2008) o Gibbs (2010), por el momento no parece que la evaluación del aprendizaje se reconozca como factor clave de la calidad institucional. Esta circunstancia alimenta una doble visión en torno al fenómeno evaluativo y acrecienta la brecha entre la teoría y la práctica que señala Zeichner (2010) para la formación universitaria.

En este contexto, las implicaciones evaluativas sujetas al EEES se sitúan en un punto intermedio entre las dos grandes visiones detectadas. Parten de directrices normativas (ENQA, 2009), pero incorporan muchas de las condiciones definidas desde la conceptualización teórica de la evaluación para el aprendizaje (Palacios & López, 2013). Esta constatación, refuerza la imagen que ofrece Zabalza (2008b) del proceso de convergencia como ocasión para introducir mejoras significativas que afecten a los sistemas de evaluación empleados en la universidad.

En suma, el modelo ITC es el producto de la integración y síntesis de las principales influencias que rodean a la praxis evaluativa en las titulaciones acotadas, tanto desde la perspectiva teórica, como desde la técnica y normativa. Serafini (2000) señala que el cambio de paradigma evaluativo debe erigirse sobre modelos de formación docente y programas de desarrollo profesional. En este sentido, la propuesta ITC representa un marco de referencia fundamentado que puede servir de soporte a este tipo de procesos.

2. OBJETIVO B. Contribuir a la superación de la ambigüedad terminológica definiendo una clasificación revisada de tipologías o modalidades evaluativas que ofrezca un marco común de referencia

A lo largo de la historia el fenómeno evaluativo ha sido interpretado desde cuatro paradigmas (Falchikov, 2005) que han ido incorporando matices y perspectivas diferentes a su definición. La creciente proliferación de estudios y publicaciones especializadas también ha contribuido a alimentar esa tendencia, dando lugar a nuevas tipologías y acepciones. Esta evolución pone de manifiesto la relevancia de la evaluación y el interés que suscita, pero también genera confusión y ambigüedad terminológica (Birenbaum, 1996; Zabalza, 2011).

La disparidad de interpretaciones que genera en torno a sí la evaluación y la necesidad de clarificación terminológica representan una constante en la literatura específica del último cuarto de siglo (B. Álvarez et al., 2007; J. M. Álvarez, 2003; Nevo, 1997; A. Pérez et al., 2007). Resulta especialmente evidente en aquellas modalidades evaluativas que tradicionalmente no han estado representadas en la educación superior, que es donde mayor confusión se acusa. Así se constata al calificar las distintas opciones dentro de la evaluación participativa o al apelar a los posibles formatos de evaluación práctica o contextualizada, donde las publicaciones especializadas evidencian la convivencia de criterios dispares.

Desde esta perspectiva, el segundo propósito de la investigación representa un prerequisite para la consecución del anterior. Planteado con carácter subsidiario, su resolución ha requerido un trabajo de síntesis y contraste que se materializa en la clasificación de tipologías evaluativas presentada en el **ANEXO 03**. Aunque parcial e incompleta, la categorización propuesta recoge las principales tipologías presentes en la literatura específica actual y ofrece un marco de referencia para favorecer la coordinación y el común entendimiento.

3. OBJETIVO C. Diseñar un modelo de competencia evaluativa para orientar procesos de formación y evaluación en el profesorado universitario dedicado a la formación básica de maestras/os

Mientras que la evaluación del aprendizaje destaca como un núcleo de investigación extensamente desarrollado, no puede afirmarse lo mismo de la competencia específica docente asociada a ella (Palacios & López, 2013). Entendida como respuesta eficaz y responsable a los dilemas y demandas de la evaluación del aprendizaje discente en el contexto formativo de referencia, la literatura específica da cuenta de escasos intentos por avanzar en su definición pormenorizada.

El tercer objetivo ha ofrecido una respuesta tentativa a tal laguna, formulando una propuesta de sistematización de la competencia evaluativa docente. Basada en la teoría de la evaluación detallada en el modelo ITC (objetivo A), su definición contempla la revisión e integración de fuentes complementarias según el proceso descrito en el capítulo quinto del marco teórico. El resultado es una propuesta que delimita ocho grandes dimensiones para la competencia evaluativa del profesorado, desarrolladas en una rúbrica que integra 19 indicadores y cinco niveles de realización en cada uno de ellos.

La definición del modelo de competencia evaluativa se ha realizado tomando como referencia el contexto particular de la formación básica de maestras/os, aunque en su elaboración se han constatado indicios que ponen en entredicho su especificidad:

- La dimensión que mejor reflejaba la adscripción de la propuesta a las titulaciones dedicadas a la preparación de maestras/os de Educación Infantil o Primaria era el modelado, relacionada con el séptimo subproceso del modelo ITC. Sin embargo su inclusión en la propuesta definitiva fue desestimada por falta de acuerdo en la fase de consulta a jueces. Sin ella, el modelo resultante pierde especificidad y se hace extrapolable a otras titulaciones universitarias con leves adecuaciones terminológicas y conceptuales.
- Los antecedentes directos consultados, refuerzan una visión general de la competencia evaluativa docente. Las pocas propuestas que han tratado de sistematizarla la abordan como una competencia docente específica y transversal, no vinculada a una etapa educativa o titulación en particular.
- El análisis de las escalas empleadas para la evaluación de la calidad docente refuerzan esa misma visión general. El estudio se ha llevado a cabo sobre

herramientas administradas al alumnado matriculado en los Grados de Educación Infantil y Primaria; sin embargo ninguna de las escalas recopiladas es específica para estas titulaciones, lo que denota que no existe el reconocimiento de un perfil docente diferenciado para las mismas (ni evaluativo, por tanto).

La consideración conjunta de las reservas presentadas ha llevado a configurar una propuesta de competencia evaluativa menos específica de lo previsto inicialmente. Sin embargo, en lo que respecta a su funcionalidad, se mantiene la vinculación con las titulaciones dedicadas a la formación básica de maestras/os, donde la competencia evaluativa del profesorado resulta especialmente relevante (C. Brew et al., 2009; V. M. López, Manrique et al., 2011; V. M. López & Palacios, 2012). De acuerdo con la descripción realizada en el capítulo tercero, en las facultades de educación y escuelas de magisterio se producen las mismas disfuncionalidades evaluativas que las constatadas en otras titulaciones, pero con riesgos adicionales. El más importante de todos ellos es su reflejo en las concepciones profundas de las próximas generaciones docentes, lo que favorecería su perpetuación y transferencia desde los niveles educativos escolares.

4. OBJETIVO D. Identificar los tópicos evaluativos presentes en las escalas de evaluación de la calidad docente administradas al alumnado en la formación básica de maestras/os (trabajo de campo intermedio)

Este cuarto objetivo ejerce un papel subsidiario con respecto al anterior en la investigación. El análisis de la presencia que tiene la evaluación del aprendizaje en las escalas administradas para la valoración de la calidad docente, representa una de las fuentes recurridas para la definición del modelo de competencia evaluativa propuesto. Éstas ofrecen la imagen latente del modelo de calidad teórica aplicable al profesorado (Ramsden, 1991; Rizo, 1999; Tejedor, 2003). En consecuencia, las alusiones que incorporen sobre la evaluación del aprendizaje pueden ser interpretadas como condiciones relevantes en este terreno e indicadores de la competencia docente.

El análisis de los tópicos evaluativos recopilados de la muestra de escalas analizada sugiere el predominio de la visión técnica y normativa de la evaluación señalada entre los referentes del modelo ITC (objetivo A). Los resultados obtenidos evidencian que los

aspectos más representados al valorar la calidad docente son los priorizados en los marcos reguladores para el ejercicio de la evaluación (ENQA, 2009; Real Decreto 1791/2010):

- La administración de información clara y transparente sobre el sistema de evaluación destaca como la condición más reiterada; a diferencia de las orientaciones teóricas de la evaluación para el aprendizaje, los ítems empleados ponen el acento en la proporción de información, no en la articulación de medidas que favorezcan su interiorización.
- El cumplimiento de los compromisos comunicados o el desarrollo coherente del dispositivo evaluativo presentado también es un tema destacado en las escalas analizadas. Parece asumido que el profesorado competente pone en práctica los dispositivos diseñados según lo previsto, sin que se haga referencia explícita a la introducción de ajustes o diversificaciones. Así lo manifiesta también Serrano (2010) indagando en el punto de vista del propio profesorado.
- Las escalas de evaluación de la calidad docente parecen participar de la preponderancia de la visión técnica e instrumental que señalan para la evaluación Knight (2005) o Verdejo (2008). Una parte destacable de los tópicos o ítems evaluativos hacen referencia a las herramientas aplicadas, en especial, a su validez y fiabilidad. La atención a estos atributos se plantea desde una perspectiva acreditativa, haciendo referencia a la capacidad de los instrumentos para ofrecer una calificación precisa de los aprendizajes previstos. No reflejan, en consecuencia, la complejidad de la que se hacen eco Mateo (2000) o De la Orden (2011b) al indagar sobre el significado de la fiabilidad y la validez cuando la evaluación no sólo sirve a las funciones sociales.

En términos generales, De Juan et al. (2007) acusan que estas escalas están diseñadas para medir habilidades típicas de la docencia tradicional y no las asociadas al planteamiento del EEES. En el terreno específico de la evaluación, así lo sugieren los resultados del análisis efectuado, que devuelven una imagen que no se corresponde con la planteada en el capítulo cuarto para el contexto formativo europeo:

- El paradigma del EEES exige que la evaluación ejerza como recurso formativo, pero la visión que reflejan las escalas analizadas es fundamentalmente acreditativa. La orientación formativa, categoría que mejor refleja la perspectiva pedagógica de la evaluación, encuentra una representación reducida entre los tópicos evaluativos.

- Para responder a su cometido, la evaluación debe trasladar la atención de los contenidos a las competencias. Sin embargo, en las escalas analizadas dicho foco de atención figura sólo de forma anecdótica.
- Unido a ello, la necesidad de desarrollar formatos de evaluación contextualizada tampoco se refleja en el banco de ítems recopilado. Aunque se haga referencia a los instrumentos y procedimientos evaluativos, el diseño de tareas de aplicación próximas a la realidad no parece vincularse a la imagen del profesorado competente.
- Suscitar la participación discente en la evaluación no representa una atribución docente prioritaria, según se desprende del análisis de las escalas consideradas. Aunque los lineamientos pedagógicos del proceso de convergencia abogan por la ampliación de los agentes evaluativos, en los tópicos recopilados sólo aparece de forma aislada.
- La apuesta por los formatos de evaluación continua, que favorezcan la concurrencia de los requisitos anteriores, tampoco se refleja de forma generalizada.

Según Villa (2008),

en el nuevo modelo europeo, el interés es comprobar lo que hace el estudiante, cómo aprende de forma autónoma y significativa, y qué resultados obtiene al final del proceso (adquisición de y logro de las competencias). Por ello, se está originando un giro a la hora de la evaluación del profesor, y se está centrando en el aprendizaje del estudiante. (pp. 189-190).

Aunque admitimos que esa sería la orientación oportuna, el análisis efectuado no permite confirmar su emergencia, al menos en materia evaluativa. Los ítems o tópicos recopilados siguen centrando la atención en la actuación evaluativa docente mayoritariamente, no en la respuesta discente que propicia la misma.

La conclusión general de la revisión planteada deja entrever una imagen de la competencia evaluativa docente que no es consistente con la elaborada en la presente investigación (objetivo C). De acuerdo con Campbell (1993/2012) son las creencias obsoletas las que dan lugar a modelos inadecuados de formación y evaluación. En consecuencia, advertimos en esta constatación un nuevo motivo para contribuir a la promoción de nuevas concepciones evaluativas más pedagógicas y consistentes con las necesidades formativas actuales.

MARCO EMPÍRICO

CAPÍTULO 7.

JUSTIFICACIÓN DEL MARCO EMPÍRICO DE LA INVESTIGACIÓN

El séptimo capítulo representa el comienzo del marco empírico de la investigación. Su desarrollo responde a la segunda pregunta central que la motiva: determinar cómo se produce en la práctica la competencia evaluativa docente, en contraste con el referente teórico construido. Para su exploración práctica, el planteamiento general de la investigación daba cuenta de cuatro objetivos que completan el camino iniciado en las fases de indagación teórica (ver tabla 57):

Tabla 57

Objetivos de la investigación abordados en el marco empírico

PROPÓSITOS DE LA INVESTIGACIÓN	
MARCO TEÓRICO	OBJETIVO ESPECÍFICO A: Formular una propuesta revisada de las principales condiciones y características que debe reunir la evaluación del aprendizaje en el contexto de la formación básica de maestras/os.
	OBJETIVO ESPECÍFICO B: Contribuir a la superación de la ambigüedad terminológica definiendo una clasificación revisada de tipologías o modalidades evaluativas que ofrezca un marco común de referencia.
	OBJETIVO ESPECÍFICO C: Diseñar un modelo de competencia evaluativa para orientar procesos de formación y evaluación en el profesorado universitario dedicado a la formación básica de maestras/os.
	OBJETIVO ESPECÍFICO D: Identificar los tópicos evaluativos presentes en las escalas de evaluación de la calidad docente administradas en las titulaciones ligadas a la formación básica de maestras/os (trabajo de campo intermedio).
MARCO EMPÍRICO	OBJETIVO ESPECÍFICO E: Realizar una autoevaluación de la competencia evaluativa del profesorado dedicado a la formación básica de maestras/os de Educación Infantil y/o Primaria en el contexto institucional de la UD y BAM.
	• OBJETIVO E1. Identificar los factores e indicadores de la evaluación del aprendizaje en los que mayor y menor dificultad expresa el profesorado, atendiendo al nivel de relevancia y realización que les atribuyen. • OBJETIVO E2. Identificar relaciones entre distintas variables académicas, socio-demográficas o profesionales y la competencia evaluativa autopercebida. • OBJETIVO E3. Identificar las principales fortalezas, debilidades y dificultades para el ejercicio de la evaluación del aprendizaje desde el punto de vista docente.

OBJETIVO ESPECÍFICO F: Realizar un diagnóstico de la competencia evaluativa del profesorado dedicado a la formación básica de maestras/os de Educación Infantil y/o Primaria en la UD y BAM desde la perspectiva de las/os estudiantes.

- **OBJETIVO F1.** Identificar los factores e indicadores de la evaluación del aprendizaje en los que mayor y menor dificultad acusa el alumnado, atendiendo al nivel de relevancia y realización que les atribuyen.
- **OBJETIVO F2.** Identificar relaciones entre distintas variables académicas, profesionales o socio-demográficas y la competencia evaluativa percibida por el alumnado.
- **OBJETIVO F3.** Identificar las principales fortalezas, debilidades y dificultades para el ejercicio de la evaluación del aprendizaje desde el punto de vista discente.

OBJETIVO ESPECÍFICO G: Contrastar el punto de vista que manifiestan docentes y discentes con respecto a la competencia evaluativa del profesorado vinculado a la formación básica de maestras/os, indagando tanto en la perspectiva teórica (importancia) como en la práctica (realización).

OBJETIVO ESPECÍFICO H: Describir las experiencias y demandas que plantea el profesorado con respecto a la formación permanente vinculada a la evaluación del aprendizaje.

- **OBJETIVO H1.** Describir los principales rasgos de las actividades formativas realizadas en los últimos años con relación a la evaluación del aprendizaje.
- **OBJETIVO H2.** Identificar las principales necesidades formativas manifestadas por el profesorado con relación al desarrollo de su competencia evaluativa.

Para determinar cómo tiene lugar la competencia evaluativa en la práctica y poder contribuir a su mejora, se ha recurrido al modelo teórico presentado en el capítulo quinto en su modalidad de uso evaluativa. Partiendo del mismo, se ha llevado a cabo un diagnóstico de la competencia evaluativa docente y de las necesidades formativas asociadas en el contexto institucional de la Universidad de Deusto (UD) y el centro adscrito Begoñako Andra Mari (BAM), ambos sitos en Vizcaya. En su desarrollo ha participado una muestra de 704 estudiantes y 51 profesoras/es vinculados a los Grados de Educación Infantil y/o Primaria.

Los próximos apartados describen y justifican el proceso de decisión que ha guiado la planificación del diagnóstico y su puesta en práctica. Tales decisiones han sido adoptadas desde una serie de supuestos y condiciones de partida que resultan indispensables para interpretar los resultados y conclusiones del estudio. En consecuencia, el nuevo capítulo se detiene en la explicitación y justificación de los siguientes aspectos en seis apartados diferentes:

- Los rasgos generales que caracterizan el diseño de investigación planteado y sus consiguientes ventajas y limitaciones.
- La presentación de la población de referencia y el procedimiento de muestreo desarrollado a partir de la misma.
- La definición de las variables sobre las que se detiene el estudio empírico.

- La presentación de los instrumentos empleados para la recogida de información en la muestra docente y discente, incluyendo su proceso de elaboración, sus propiedades psicométricas, así como las posibilidades y limitaciones que ofrecen.
- El procedimiento de recogida de información articulado.
- Los principales análisis efectuados a partir de los datos e informaciones recabadas.

Como en los capítulos anteriores, este recorrido se completa con una síntesis final de las principales cuestiones abordadas en su desarrollo.

1. Diseño del marco empírico de la investigación

La consecución de los propósitos formulados exige el establecimiento de un diseño de investigación que favorezca su resolución. “El término diseño hace referencia a la elaboración de un plan de actuación una vez que se ha establecido el problema de investigación” (León & Montero, 2012, p. 14) y se caracteriza por la elección consciente y deliberada de los procedimientos, métodos e instrumentos adecuados para su resolución. Conduce al establecimiento de un plan de trabajo ligado a los objetivos perseguidos y a los presupuestos ontológicos, epistemológicos y metodológicos desde los que se articulan. Albert (2009) lo sintetiza en los siguientes términos:

El término diseño se refiere al plan o estrategia concebida para obtener la información que se desea. Los diseños de investigación se inventaron para permitir a los investigadores responder preguntas de forma más válida, objetiva, precisa, y económica posible. Los planes de investigación se conciben de forma deliberada y específica y son ejecutados para obtener evidencia empírica que apoye al problema de investigación. (p. 58).

En términos generales, el diseño responde a una doble finalidad: procurar que las preguntas que motivan la investigación queden resueltas y controlar la influencia que puedan ocasionar factores desconocidos o ajenos en ese proceso. Para ello, exige la adopción de decisiones en niveles complementarios que afectan a la propia estructura del

trabajo de campo, su emplazamiento físico y ordenación cronológica, los procedimientos de muestreo empleados o la instrumentación articulada para la recogida de datos y evidencias.

La consideración de los objetivos y el alcance de la investigación hacen que cada diseño adopte una serie de rasgos propios, dentro de un rango de variedad muy amplio:

- El diseño de la presente investigación plantea un estudio *no experimental* y *descriptivo* sobre la competencia evaluativa docente, enfocado a comprender cómo tiene lugar la misma en un contexto formativo determinado: las titulaciones vinculadas a la formación básica de maestras/os en la UD y BAM. También adopta ciertos matices que le aproximan a los estudios *comparativo-causales* y *correlacionales*; dentro de los límites que determina la no manipulación de las variables independientes, trata de explicar las diferentes percepciones que se constatan en torno a la competencia evaluativa y explora posibles relaciones de asociación en torno a ella. En consecuencia, atendiendo a su finalidad, el diseño de investigación configurado integra *rasgos descriptivos y explicativos*.
- Frente a las variables independientes, que están determinadas antes de la investigación (*ex post facto*), el desenlace o manifestación de la competencia evaluativa (variable dependiente) tiene lugar en el presente. Por este motivo, el diseño de la investigación es *descriptivo y prospectivo* atendiendo a su cronología y su sentido temporal.
- Considerando la ubicación o emplazamiento, el estudio diseñado representa un *trabajo de campo*, desarrollado en el mismo medio en el que habitualmente se manifiesta la competencia evaluativa del profesorado. Se trata de una *investigación participativa*, desarrollada en respuesta a una inquietud compartida y en interacción las/os principales afectadas/os.
- El proceso formal de investigación planteado se ajusta al esquema de desarrollo de los *estudios inductivos*. Por la complejidad y singularidad del objeto de estudio que atiende, centra la atención en el análisis de casos particulares para llegar a extraer conclusiones generales dentro de los límites que posibilita la muestra. Tomada de un contexto institucional específico, esta última da lugar a un *estudio de caso*. En consecuencia, las descripciones y explicaciones generadas sólo son aplicables en las condiciones y circunstancias particulares que concurren en el mismo.
- En cuanto a la dimensión temporal, la investigación ha sido planteada de forma *transversal o transeccional*, contando con la participación de docentes y estudiantes de los distintos segmentos de edad de forma simultánea. Su orientación es prioritariamente *descriptiva*, pero alberga ciertas pretensiones *exploratorias*. Los

estudios descriptivos pretenden categorizar y ofrecer una visión de un determinado contexto, fenómeno, comunidad o situación para describirla (Albert, 2009). El establecimiento de causas y correlaciones generalizables está limitado por el tipo de muestra que ha sido configurada (estudio de caso); sin embargo, este análisis puede ser concebido como una exploración tentativa del modelo teórico ideado, para favorecer aproximaciones más profundas y extensas en futuras investigaciones.

- Por último, en cuanto al criterio de datación o la naturaleza de los datos recopilados y analizados, la presente investigación incorpora elementos cualitativos y cuantificables aproximándose a la *convivencia pluri-metodológica* (Sabariego & Bisquerra, 2009a). La combinación de la metodología de investigación cualitativa y cuantitativa sitúa este proyecto en una posición intermedia entre las fortalezas y debilidades de una y otra, aunque ambos planteamientos no están representados por igual en el diseño de la investigación.

No existe un diseño perfecto, especialmente cuando la indagación apunta a un contexto complejo como es el educativo. A pesar de constituir el resultado de un ejercicio de elección consciente y coherente, este particular diseño de investigación supone asumir una serie de ventajas y limitaciones que condicionan el proceso completo de estudio empírico y la calidad de sus conclusiones. Por este motivo, la tabla 58 sintetiza las principales ventajas y limitaciones derivadas de sus rasgos y contribuye al establecimiento de expectativas ajustadas sobre las posibilidades y limitaciones de la propuesta de análisis y sus resultados.

Tabla 58

Ventajas y limitaciones de los rasgos del diseño de investigación ideado

RASGO	IMPLICACIONES: VENTAJAS Y LIMITACIONES ASOCIADAS	
DISEÑO NO EXPERIMENTAL Y DE CAMPO	VENTAJAS	• Potencialmente, mejor ajuste que los diseños experimentales para valorar las percepciones de docentes y discentes sobre una realidad compleja como es la competencia evaluativa y las necesidades formativas vinculadas a la misma. No es posible aislar todos los factores que afectan tal percepción, ni mucho menos manipularlos, por lo que constituye la única alternativa factible. • Realiza el estudio en las circunstancias naturales en las que tiene lugar la evaluación del aprendizaje, minimizando la artificialidad y el impacto de variables extrañas asociadas al propio proceso de investigación. • Unido a ello, se produce una mayor validez externa, ya que las condiciones en las que ha sido llevado a cabo el estudio se asemejan a la realidad mucho más que una situación recreada o controlada artificialmente.

RASGO		IMPLICACIONES: VENTAJAS Y LIMITACIONES ASOCIADAS
DISEÑO NO EXPERIMENTAL Y DE CAMPO	LIMITACIONES	• Imposibilidad de manipulación de las variables independientes y, en consecuencia, carencia de control. • Unido a lo anterior, riesgo de interpretaciones inadecuadas: ✓ No se pueden establecer relaciones causales entre las variables analizadas. ✓ Grandes dificultades para generalizar los resultados y extraer conclusiones prospectivas. • Se reduce la validez interna, ya que en el fenómeno estudiado pueden intervenir variables extrañas no consideradas o producirse relaciones no identificadas. Además, las posibilidades de réplica son muy limitadas.
DESCRIPTIVO (por finalidad y criterio cronológico) Y PROSPECTIVO	VENTAJAS	• Evita los sesgos de imprecisión ligados a la retrospección. • Ofrece una descripción completa del fenómeno estudiado tal y como se vivencia en un grupo, momento y situación particular.
	LIMITACIONES	• No da lugar a explicaciones causales ni permite plantear pronósticos a futuro.
ESTUDIO DE CASO	VENTAJAS	• Permite una exploración contextualizada, sensible a las circunstancias específicas. Riqueza de datos. • Accesibilidad, viabilidad del proyecto. • Puede ofrecer una representación simplificada útil para encauzar estudios de grupo en el futuro.
	LIMITACIONES	• Imposibilidad de generalización de resultados más allá del grupo que conforma la población de referencia. No hay representatividad general. • Asociada a dicha imposibilidad, escasa rentabilidad e impacto de la investigación. • Falta de aleatorización en la muestra, al no poder garantizar un mismo nivel de representación de todas las variables independientes consideradas. • No es posible garantizar en la muestra los criterios de normalidad y homocedasticidad, lo que condiciona la realización de ciertos análisis.
INDUCTIVO	VENTAJAS	• Potencialmente, este tipo de diseño se adapta mejor a la complejidad y la singularidad de los contextos educativos, donde las normas generales a menudo no resultan directamente aplicables. • Favorece una descripción profunda y completa de un grupo, contexto o comunidad determinada. Conocimiento idiográfico.
	LIMITACIONES	• Dificultades para generalizar las conclusiones extraídas a otras situaciones.
TRANSVERSAL	VENTAJAS	• Resulta más económico y viable que un proyecto longitudinal.
	LIMITACIONES	• No permite profundizar en el desarrollo real que tiene la competencia evaluativa percibida a lo largo de la carrera docente y discente. • Las covarianzas sistemáticas en variables como la edad o el curso, no pueden interpretarse directamente en términos causales al estar implicadas otras variables. • Dificultades para configurar muestras representativas.

RASGO		IMPLICACIONES: VENTAJAS Y LIMITACIONES ASOCIADAS
COMBINACIÓN DE DATOS CUALITATIVOS Y CUANTITATIVOS	VENTAJAS	- Economía. Permite un acceso a la información más amplio y rápido y el análisis es también más sencillo y menos laborioso. - Permite complementar las perspectivas procedentes de los dos enfoques metodológicos equilibrando fortalezas y debilidades.
	LIMITACIONES	Preponderancia de la información cuantitativa sobre la cualitativa, con los riesgos asociados que ello implica (superficialidad, artificialidad, etc.).
PARTICIPATIVA	VENTAJAS	- Sensibilidad con el objeto de estudio y conocimiento previo, lo que aumenta la probabilidad de un diseño más adecuado. - Mayor facilidad de acceso a la muestra. - Puede hacer de la investigación una herramienta formativa para la reflexión y la mejora.
	LIMITACIONES	- Pérdida de subjetividad del investigador, que es además agente implicado en el fenómeno en estudio. - Confusión o falta de transparencia en la delimitación de roles (investigador vs. docente).

Nota: Reelaboración propia a partir de las aportaciones de Albert (2009), León y Montero (2012) y Torrado (2009).

2. Población y muestra

La determinación de la muestra constituye una de las decisiones clave que se debe adoptar en el diseño de la investigación. La posibilidad de generalización o transferencia de las conclusiones alcanzadas y el sentido que adopten las mismas dependen de la cantidad de sujetos que se hayan consultado, de los procedimientos aplicados para su selección y de su capacidad para representar al colectivo de referencia.

La primera delimitación que se debe realizar en este sentido apunta al colectivo hipotético (Torrado, 2009) o **universo** al que se orienta el estudio. Éste lo configuran todos los posibles sujetos de cierto tipo (Latorre, Del Rincón & Arnal, 1996) lo que, en la investigación en curso, remite al alumnado y profesorado vinculado a la formación básica de maestras/os de Educación Infantil y Primaria en todo el estado.

La **población**, entendida como el conjunto de individuos en los que se desea estudiar el fenómeno (Albert, 2009), acota la delimitación inicial del universo. En nuestro caso centra la atención en el contexto institucional de la UD y el centro adscrito BAM, reduciendo la participación potencial a las/os 67 docentes y 1693 estudiantes vinculados a

los grados reseñados. Resultaría de indudable interés la inclusión de otros centros vinculados a la formación básica de maestras/os hasta cerrar la población a nivel autonómico o estatal. Sin embargo, un primer sondeo en Navarra y la Comunidad Autónoma del País Vasco nos hizo valorar serias dificultades desde el punto de vista de la accesibilidad y la viabilidad.

La imposibilidad de trabajar con la población al completo, ha derivado en la selección de una **muestra** de 704 estudiantes y 51 profesoras/es, configurada en dos fases. Su definición se ha efectuado de acuerdo a los principios de control y economía (Manzano & Braña, 2005), valorando los propósitos y el alcance del estudio, así como las condiciones prácticas asociadas a su viabilidad (Sabariego, 2009).

La muestra docente y discente ha sido conformada de acuerdo al modelo de los ciclos de muestreo de Fox (1981). Este nuevo apartado presenta las principales decisiones adoptadas en cada una de las cinco fases, desarrollando el recorrido que sintetiza la figura 25 y que describe en profundidad el **ANEXO 12**:



Figura 25. Resumen del proceso de muestreo desarrollado según Fox (1981).

Los dos próximos apartados describen la representatividad y el tamaño de la muestra resultante, aptas para llevar a cabo el estudio previsto.

2.1 Representatividad de la muestra.

La representatividad de la muestra condiciona la generalización de los resultados obtenidos a la población y, con ello, el impacto o rendimiento de la investigación. Se satisface cuando “trabajar con ella nos permita llegar a las mismas conclusiones a las que habríamos llegado de haber trabajado directamente con la población, si ello hubiera resultado viable o aconsejable” (Manzano & Braña, 2005, p. 95).

Concebida como estudio de caso, la transferencia y generalización de los resultados se encuentra fuera de las pretensiones de la investigación. En consecuencia, se ha recurrido a un procedimiento de muestreo no probabilístico casual o accidental, configurado por cuotas:

- El muestreo no probabilístico realizado es casual o accidental en atención a dos circunstancias relevantes. Por una parte, su selección parte de una disposición inicial de accesibilidad, representando lo que Stake (1995) denomina una oportunidad para aprender. Por la otra, la selección de los casos de la muestra ha estado determinada por la aceptación voluntaria de participación en un momento determinado del curso. Por efecto de la accidentalidad, ha quedado fuera de la muestra docente el profesorado no dispuesto a participar, así como aquel otro que no haya podido recibir la oferta (bajas, excedencias, estancias en el extranjero, etc.) o que no tuviera docencia adjudicada en el momento de la solicitud.

La accidentalidad es aún más evidente en la configuración de la muestra de estudiantes, que se ha realizado por conglomerados, utilizando la agrupación académica de referencia. Su selección inicial depende directamente de la muestra docente. Cada profesor/a participante, aporta un grupo de estudiantes a la muestra potencial, siendo éste elegido por el equipo investigador de acuerdo con un sistema de cuotas expuesto a continuación. En consecuencia, las/os integrantes definitivas/os de la muestra discente son aquellas/os adscritos al grupo seleccionado, presentes en el aula el día de la recogida de información y dispuestos a colaborar en la misma (ver figura 26).

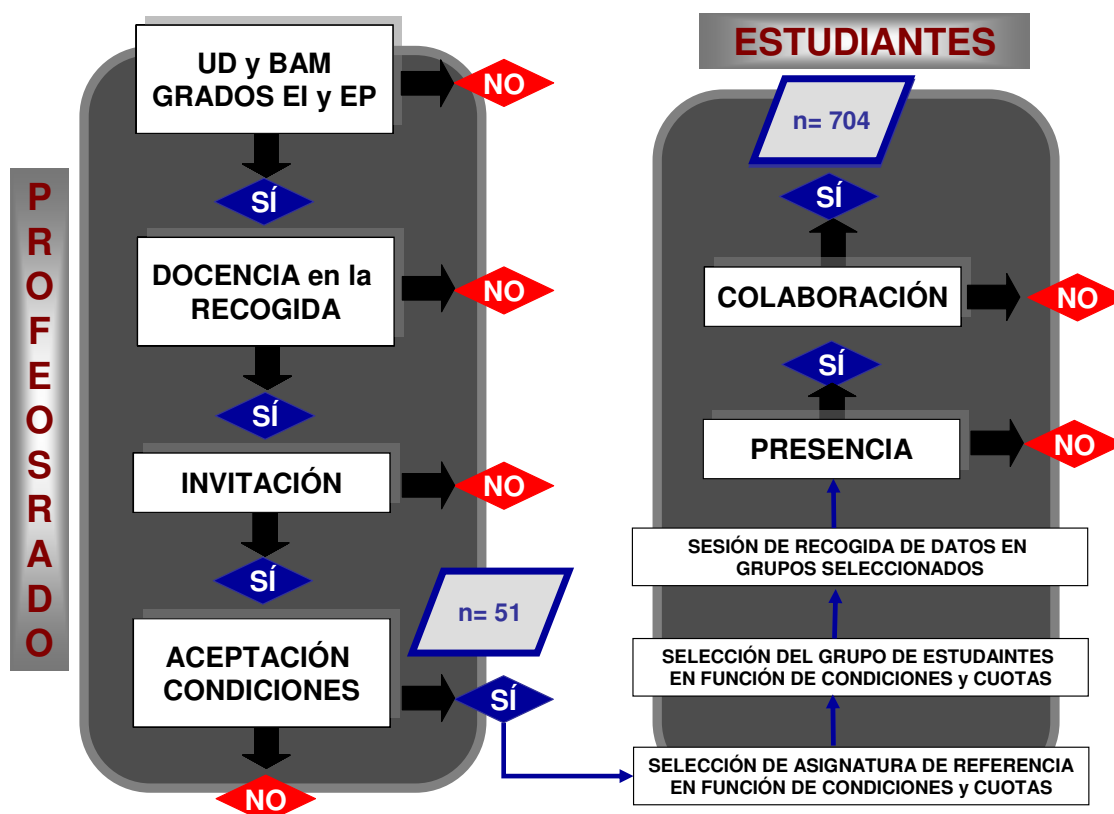


Figura 26. La incidencia de la accidentalidad en la muestra docente y discente.

- A pesar de que el muestreo efectuado no responda al criterio de equiprobabilidad, la selección de casos ha sido efectuada de acuerdo a una serie de cuotas presentes en la población de referencia. Aunque el criterio de accesibilidad es prioritario, se ha procurado la representación ajustada de tres estratos: centro (UD y BAM), titulación (Educación Infantil y Educación Primaria) y curso (1º, 2º ó 3º, ya que en el último curso no se hallaba implantado en ninguno de los centros al iniciar la recogida).

En suma, la muestra de estudiantes conformada representa el 41,58% de la población discente total. La tabla 59 especifica el nivel de representatividad logrado por cuotas o estratos, indicando en cada celda el número de casos registrados en la población (N), el número de casos incluidos en la muestra (n) y la proporción que suponen. Asumido el curso como estrato básico para la configuración de la muestra, la representación media obtenida por cuota es del 42,47% (42,95% en la primera fase y 41,98% en la segunda). Las principales discrepancias constatadas respecto a la misma se asocian a la accesibilidad y a la propia accidentalidad del muestreo.

Tabla 59

Síntesis de la representatividad de la población en la muestra discente, según datos facilitados por la Secretaría Académica de la Facultad de Psicología y Educación de la UD y de BAM

		BAM			BAM			UD				
		GRADO EDUCACIÓN INFANTIL			GRADO EDUCACIÓN PRIMARIA			GRADO EDUCACIÓN PRIMARIA				
		1º	2º	3º	1º	2º	3º	1º	2º	3º	TOTAL	
FASE I	2011/2012	70	74	58	140	133	114	94	83	68	834	
		32 (45,7%)	26 (35,1%)	31 (53,4%)	64 ^a (45,7%)	45 (33,8%)	40 (35,0%)	41 (45,0%)	0 ^b (0%)	63 (92,6%)	342 (40,0%)	
FASE II	2012/2013	71	65	73	120	135	141	78	93	83	859	
		58 (81,6%)	0 ^c (0%)	16 ^d (21,9%)	91 (75,8%)	52 (38,5%)	41 (29,0%)	72 (94,3%)	0 ^c (0%)	32 (38,5%)	362 (42,1%)	
											N	1693
											n	704
											Representación de la población en la muestra	41,5%

Representación de la población en la muestra

Nota: La división superior de las celdas, sombreada en amarillo, detalla los datos relativos a la población. La parte central de la celda indica el número de casos integrados en la muestra y el porcentaje de representación que suponen.

^a En este mismo grupo se realizó una aplicación piloto sobre 29 estudiantes que no ha sido computada dentro de la muestra. El número de casos reales en esta celda descendería a 111, lo que supondría que la representación real obtenida es 57,65%.

^b La consulta no es posible por problemas de agenda (**ANEXO 12**).

^c No hay muestra disponible porque el conglomerado fue consultado en la Fase I (1º en 2011/2012). No es posible discriminar dentro del grupo el alumnado que no participó en la primera recogida sin vulnerar su anonimato.

^d Sólo queda un/a docente pendiente de consulta, por lo que sólo se puede acceder a un grupo de estudiantes (**ANEXO 12**).

En síntesis, la muestra discente queda conformada por 496 estudiantes de BAM (70,5%) y 208 más matriculadas/os en el campus bilbaíno de la UD (el 29,5% restante). Las/os estudiantes del Grado de Educación Infantil, representan el 23,2% de la muestra y proceden en su totalidad del centro adscrito⁷⁹. El alumnado matriculado en el Grado de Educación Primaria representa el 76,8% y se reparte entre ambas instituciones: el 47,3% realiza sus estudios en BAM y el 29,5% en la UD. Dentro de los límites declarados, se ha procurado la representación de los tres primeros cursos de las titulaciones, quedando configurada la muestra definitiva como refleja la figura 27:

⁷⁹ Esta titulación no se imparte en la UD.

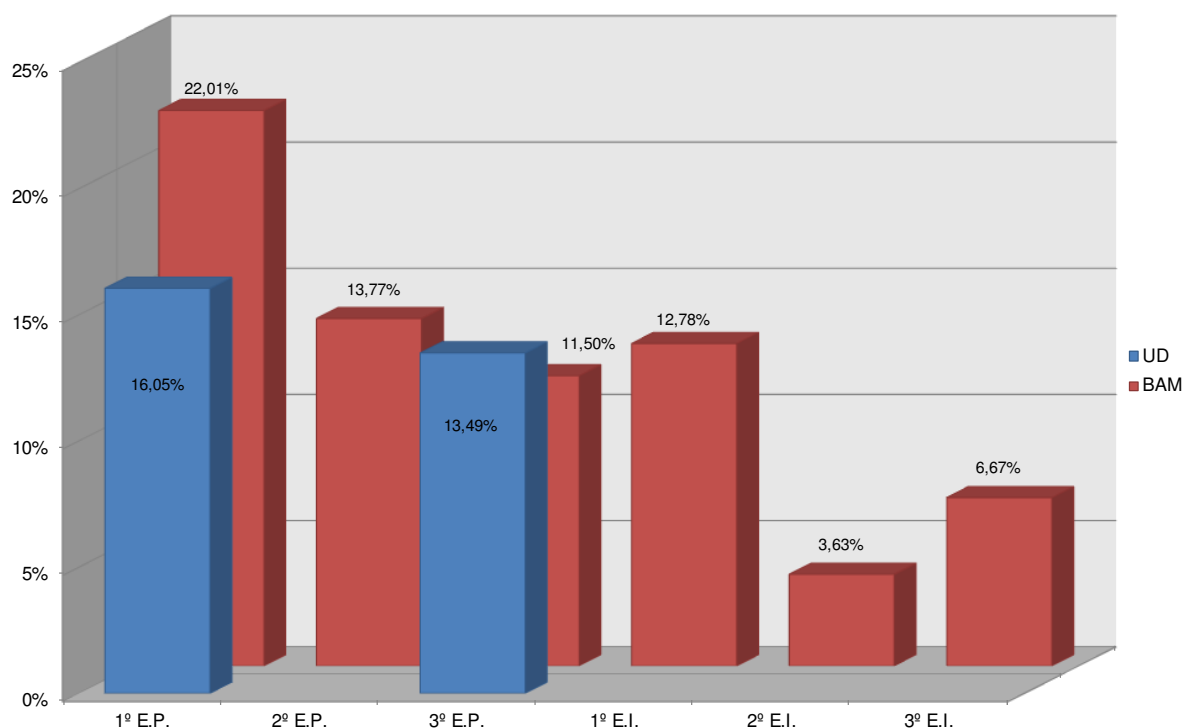


Figura 27. Configuración de la muestra discente por estratos: centro (BAM y UD), titulación (EI- Educación Infantil y EP- Educación Primaria) y curso (1º, 2º y 3º).

En cuanto a la **muestra docente**, representa el 76,11% de la población de referencia concentrando la participación de 51 sujetos (17 de la UD y 34 de BAM). En ambos centros, el profesorado ejerce simultáneamente en varios cursos y/o en ambas titulaciones, de modo que no es posible ofrecer un mapa claro de la representación poblacional.

La población de profesoras/es en el Grado de Educación Primaria de la UD asciende a 27 y se ha reunido en muestra al 65,38% (17 casos, 33,33% de la muestra docente conformada). La población docente en BAM está conformada por 40 sujetos, en su mayoría implicados en las dos titulaciones ofertadas: en el Grado de Educación Infantil el encargo docente lo desempeñan 23 profesoras/es y en el Educación Primaria 39. La muestra reúne al 85% del claustro, con 12 representantes del equipo docente de la primera titulación (23,52% de la muestra total) y otros 22 de la segunda (43,13% de la muestra docente resultante). La figura 28 detalla la afijación de casos en este sector de la muestra por centro, titulación y curso:

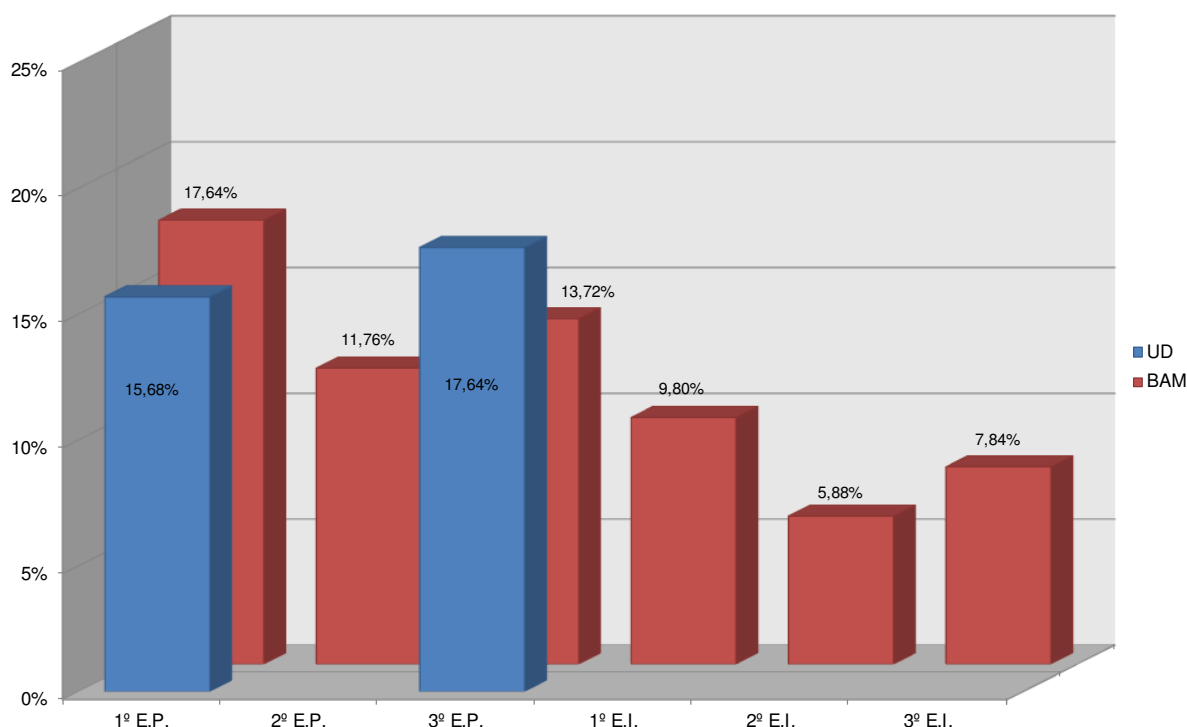


Figura 28. Configuración de la muestra docente por estratos: centro (BAM y UD), titulación (EI- Educación Infantil y EP- Educación Primaria) y curso (1º, 2º y 3º).

2.2 Tamaño de la muestra.

Los propósitos de la investigación no plantean el establecimiento de predicciones y generalizaciones, en consecuencia, el tamaño de la muestra se ha determinado sin recurrir a procedimientos estadísticos. Sin embargo, para favorecer su calidad y suficiencia, se han establecido una serie de mínimos partiendo de las propias exigencias de la investigación:

- En ambos sectores se ha establecido como requisito de partida la configuración de *muestras superiores a 30 individuos*, tamaño mínimo requerido para la realización de algunos de los análisis inferenciales más frecuentes (R. Álvarez, 1995; Berlanga & Rubio, 2012; Iglesias & Sulé, 2005; Vilá & Bisquerra, 2009).
- También se ha considerado el *tamaño de la población de referencia* como criterio para determinar la cantidad mínima de casos a incorporar en ambas muestras. Bajo la premisa de que a mayor tamaño de población, menor porcentaje de representación se requiere, Cardona (2002) ofrece una serie de orientaciones generales a este respecto. Como recoge la tabla 60, la muestra discente supera ampliamente la estimación del 20% que correspondería a la magnitud de su

población de origen. No así la muestra docente, donde teóricamente se debería haber accedido al 100% de los casos por proceder de una población pequeña ($N = 67$). Anticipando esta dificultad por las condiciones del muestreo y la voluntariedad de la experiencia, tal estimación ha sido matizada a priori para la presente investigación con un cupo mínimo de acceso al 50% del profesorado de la población, condición que sí se satisface en la muestra conformada.

Tabla 60
Población (N) y tamaño de la muestra (n) según orientaciones de Carmona (2002)

ESTUDIANTES		
POBLACIÓN (N)	TAMAÑO n SUGERIDA	MUESTRA (n)
100 sujetos o menos	Población completa ($n = N$)	-
500 sujetos aproximadamente	50% ($n = 250$ aproximadamente)	-
1500 sujetos aproximadamente ($N = 1693$)	20% ($n = 300$ aproximadamente)	41,58% (704)
Más de 5000 sujetos	$n = 400$	
DOCENTES		
POBLACIÓN (N)	TAMAÑO n SUGERIDA	MUESTRA (n)
100 sujetos o menos ($N = 67$)	Población completa ($n = N$)	NO
	Mínimo 50% (criterio interno)	76,11% (51)

3. Definición de variables

Entendidas como características o propiedades que adoptan valores diferentes entre los sujetos (Sabariego, 2009), la determinación de las variables está estrechamente ligada a los propósitos que persigue la investigación.

Existen diferentes criterios para ordenar los distintos tipos de variables. En función de la relación que guardan con los propósitos o hipótesis analizadas, se diferencian variables independientes, dependientes y extrañas. La tabla 61 ordena las variables consideradas en el análisis de la competencia evaluativa docente de acuerdo a dicho criterio. Como en ella se advierte, algunas variables han sido consideradas en ambas muestras (docente y discente) mientras que otras responden específicamente a una de ellas:

Tabla 61

Clasificación de las variables contempladas en el marco de la investigación

VARIABLES INDEPENDIENTES	COMUNES	• Demográficas: ✓ Fuente. ✓ Sexo. ✓ Edad. • Académico-institucionales: ✓ Módulo/Asignatura valorada/impartida. ✓ Unidad (módulo o asignatura). ✓ Titulación en la que se incluye. ✓ Curso. ✓ Centro. ✓ Módulo. ✓ Materia.
	DOCENTES	• Académico-profesionales: ✓ Experiencia docente universitaria. ✓ Formación profesional previa: titulación inicial de acceso. ✓ Titulación postgrado: máster. ✓ Titulación postgrado: doctorado. ✓ Tipo de dedicación a la docencia. ✓ Simultaneidad de la docencia con otra ocupación educativa. • Influencias posibles vinculadas a la competencia evaluativa: ✓ Autopercepción general de competencia evaluativa (tres niveles). ✓ Formación permanente en evaluación (realización: sí o no). ✓ Necesidad percibida de formación en evaluación (sí o no).
	DISCENTES	• Académico-profesionales: ✓ Vía de acceso a la titulación. ✓ Dedicación a la formación. • Influencias posibles: ✓ Experiencia previa con la/el docente de referencia. ✓ Calificaciones/resultados parciales recibidos. ✓ Expectativa de logro en la asignatura cursada. ✓ Percepción general de competencia docente. ✓ Modalidad de aplicación del cuestionario.
VARIABLES DEPENDIENTES	DOCENTES	• Relevancia atribuida a la competencia evaluativa (total y por dimensiones e indicadores). • Competencia evaluativa autopercebida (total y por dimensiones e indicadores). • Formación permanente realizada en evaluación (abierta con ciertas indicaciones): ✓ Voluntariedad. ✓ Duración. ✓ Temas. ✓ Modalidad. ✓ Satisfacción. • Formación permanente demandada en evaluación (abierta). • Fortalezas y áreas de mejora en la evaluación del aprendizaje (abierta).

	DISCENTES	• Relevancia atribuida a la competencia evaluativa (total y por dimensiones e indicadores). • Competencia evaluativa percibida (total y por dimensiones e indicadores). • Fortalezas y áreas de mejora en la evaluación del aprendizaje (abierta).

3.1 Variables dependientes.

Desde la perspectiva de la investigación en curso, las variables dependientes son aquellos fenómenos o conductas cuya descripción y mejora se pretende. Como se expone en el próximo apartado, se suponen afectadas por el influjo de los factores o variables independientes.

Como recoge la tabla 61, las variables dependientes consideradas en la presente investigación giran en torno a su principal objeto de estudio, a saber, la competencia evaluativa del profesorado vinculado a los Grados de Educación Infantil y Primaria en los centros de referencia. Esta competencia evaluativa representa un constructo (León & Montero, 2012) o variable latente (Sabariego, 2009), que no se presta a cuantificación directa. Para favorecer su análisis, se presenta operativizada por medio de las dimensiones, indicadores y descriptores delimitados en su sistematización en el capítulo quinto. Ante la imposibilidad de efectuar comprobaciones directas, por cada uno de los indicadores considerados se recogen dos datos mensurables que proporcionan una referencia indirecta:

- El grado **relevancia o importancia** que atribuye el sujeto informante (docente o discente) a cada uno de los indicadores de competencia evaluativa. Por cada indicador, se conforma una variable cuantitativa discreta con una escala de cinco valores, ordenados de menor a mayor importancia: 1 (*irrelevante*), 2 (*poco relevante*), 3 (*relevancia moderada*), 4 (*relevante*) y 5 (*muy relevante*). Su medición se realiza por medio de los cuestionarios expuestos en el próximo apartado para cada fuente.
- El grado de **cumplimiento o realización** que percibe el sujeto informante (docente o discente) en cada uno de los indicadores de competencia evaluativa. Como en el caso anterior, se concreta en una variable cuantitativa discreta, sujeta a una escala con cinco posibles valores: 1 (*inadecuado*), 2 (*poco adecuado*), 3 (*adecuado*), 4 (*muy adecuado*) y 5 (*excelente*). Su medición también se realiza por medio de

cuestionario en ambas fuentes, en un ejercicio de autoevaluación o evaluación, según se oriente a alumnado o profesorado.

Además de la competencia evaluativa docente, la investigación plantea entre sus propósitos la indagación sobre las siguientes cuestiones:

- **Fortalezas y área de mejora en la evaluación del aprendizaje.** Se trata de una variable cualitativa recogida como pregunta abierta en los cuestionarios de estudiantes y personal docente.
- **Formación permanente realizada en evaluación durante los últimos dos años.** Exclusivamente dirigida al profesorado, es otra variable cualitativa centrada en la descripción de las actividades formativas desarrolladas recientemente sobre la evaluación del aprendizaje. Aparece planteada como pregunta abierta, aunque con indicaciones para la construcción de la respuesta o descripción (duración, voluntariedad, modalidad, temas y satisfacción).
- **Formación permanente demandada en evaluación.** Es otra variable cualitativa específicamente orientada al colectivo docente. En una cuestión abierta, invita al profesorado a describir el tipo de formación en evaluación que precisa o considera conveniente.

Como muestra el próximo capítulo, las dos últimas variables, relativas a la formación, también han sido empleadas como factor independiente para el desarrollo de estudios inferenciales. En ese caso, su empleo ha permitido diferenciar dentro de la muestra docente dos grupos: profesorado con y sin formación permanente o con y sin percepción de necesidad formativa.

3.2 Variables independientes.

En la presente investigación las variables independientes son aquellos factores que podrían ocasionar algún efecto en la competencia evaluativa docente percibida o autopercebida y en las necesidades formativas vinculadas a la misma. Hacen referencia a aspectos de naturaleza diversa con influencia potencial en las variables dependientes definidas. En función de su origen y carácter, pueden clasificarse como variables

demográficas, académico-institucionales, académico profesionales o como posibles fuentes de influencia.

3.2.1 Variables demográficas.

Como su propio nombre indica, hacen referencia a factores de carácter demográfico fácilmente objetivables en ambas muestras. Son tres y aparecen entre los datos de identificación de las herramientas de recogida diseñadas:

- **Fuente.** Es una variable cualitativa dicotómica, sujeta a una escala nominal con dos posibles valores (docente y estudiante).
- **Sexo.** Variable cualitativa dicotómica relativa al género de la persona informante. Responde a una escala nominal con las opciones clásicas de hombre y mujer.
- **Edad.** Se trata de una variable cuantitativa recogida como valor discreto y posteriormente recodificada por intervalos.

3.2.2 Variables académico-institucionales.

Son datos objetivos que hacen referencia al centro, el plan de estudios y al nivel desde el que se efectúa la valoración de la competencia evaluativa (asignatura/módulo y curso). Son siete variables compartidas para estudiantes y profesorado:

- **Asignatura/módulo valorada/impartida.** Se trata de una variable cualitativa *politómica*, sujeta a medición en escala nominal. Establece el título del módulo o asignatura en la que se centra el análisis de la competencia evaluativa, según haya sido determinado en el procedimiento de muestreo (a propuesta del equipo investigador y en consenso con el profesorado participante como detalla el **ANEXO 12**). Su registro ha dado lugar a un total de 32 categorías enumeradas en el **ANEXO 13**.
- **Unidad docente de referencia.** Representa una variable cualitativa dicotómica que discrimina las valoraciones contextualizadas en un módulo (3 ECTS) o aquellas otras realizadas para asignaturas (6 ECTS).
- **Titulación.** La titulación indica si el módulo o asignatura en la que se contextualiza la valoración se adscribe al Grado de Educación Infantil o al de Educación Primaria. Es una variable cualitativa dicotómica de escala nominal.

- **Curso.** Hace referencia al año de la titulación en la que se imparte el módulo/asignatura en la que se realiza la valoración de la competencia evaluativa docente. Es una variable cualitativa *politómica* de escala ordinal que adopta tres posibles valores (1º, 2º ó 3º), según las precisiones efectuadas en la población.
- **Centro.** Se trata de una variable cualitativa dicotómica de escala nominal. Indica en cuál de los dos centros se lleva a cabo el análisis (campus bilbaíno de la UD o centro adscrito BAM).
- **Módulo.** Se trata de una variable cualitativa *politómica* que detalla el tipo de módulo al que se adscribe la unidad docente de referencia, según los decretos reguladores de las titulaciones conducentes a los títulos de Maestra/o en Educación Infantil y Primaria (las órdenes ECI 3854/2007 y ECI 3857/2007, respectivamente, ambas dependientes de lo establecido en el Real Decreto 1393/2007 para la ordenación de las enseñanzas universitarias). Con acuerdo a los mismos, esta variable incorpora cuatro categorías de respuesta: formación básica, módulo didáctico-disciplinar, de mención o DECA (Declaración Eclesiástica de Competencia Académica)⁸⁰.
- **Materia.** Igual que el módulo de referencia, esta variable cualitativa *politómica* depende de la adscripción de las unidades de valoración a las materias establecidas en los decretos reguladores reseñados. En el plan de estudios conducente al título de Maestra/o en Educación Primaria se discriminan 14 materias, dos más que las presentes en el Grado de Educación Infantil (**ANEXO 13**).

3.2.3 Variables académico-profesionales.

Como indica su nombre, hacen referencia a diversos factores de naturaleza académica y profesional. Son diferentes para estudiantes y profesorado y así se presentan bajo estas líneas.

En el caso del **profesorado**, han sido consideradas seis variables de esta naturaleza:

- **Experiencia docente universitaria.** Presentada como variable cuantitativa continua, recoge el número de años de dedicación a la docencia universitaria en el

⁸⁰ Este último módulo representa una aportación institucional al plan de estudios y que está sujeta a una acreditación opcional. En este listado falta el módulo Practicum, que ha sido excluido de los contextos de valoración en atención a sus particularidades organizativas y evaluativas.

momento de la recogida de datos. Esta variable ha sido recodificada por intervalos en la fase de análisis.

- **Formación profesional previa o titulación inicial de acceso.** Establece la titulación académica o formación profesional básica de grado o equivalente. Su denominación parte de la distinción que propone Varcárcel (2003) para las fases de la formación docente (previa, inicial, continua y especialización didáctico-disciplinar), aunque ha sido matizada para su mejor ajuste a los propósitos de la presente investigación. Este dato ha sido recabado como cuestión abierta dando lugar a 10 categorías diferentes de formación previa.
- **Formación de postgrado a nivel de máster.** Se trata de una variable cualitativa *politómica* medida en escala nominal y referida a la formación de postgrado a nivel de máster. Establece tres categorías para su análisis: sí (con titulación de máster), no (hasta la fecha, sin estudios a ese nivel) y en proceso.
- **Formación de postgrado a nivel de doctorado.** Esta nueva variable se plantea en paralelo a la anterior, pero atendiendo a la formación de postgrado al nivel de doctorado.
- **Tipo de dedicación.** Hace referencia al tipo de contrato que vincula a la persona informante con el ejercicio de la docencia universitaria. Para su definición, ha sido adoptado el sistema de reconocimiento empleado en el centro adscrito BAM, por ser el que mayor número de casos proporciona a la muestra. En función del número de horas de contrato, el mismo establece que la dedicación puede ser exclusiva (37,5 horas semanales), plena (30 horas semanales) o parcial (menos de 30 horas de dedicación semanal). Es una variable cualitativa *politómica* ordinal.
- **Simultaneidad de la docencia con alguna otra actividad educativa regulada.** Representa un factor cualitativo dicotómico que recoge la dedicación exclusiva o no del profesorado consultado a la docencia universitaria. Para ello, se plantea a las/os participantes si, además de la dedicación declarada a la enseñanza universitaria, desempeñan alguna otra función docente en otras instituciones educativas (universitarias o no universitarias).

En la **muestra de estudiantes** se han registrado dos factores de índole académico-profesional:

- **Vía de acceso a la titulación.** Consiste en una variable cualitativa *politómica* que determina, en escala nominal, el itinerario desde el cual ha accedido el alumnado al Grado de Educación Infantil o Primaria. Está sujeta a cinco posibles alternativas de

respuesta: prueba de acceso desde Bachillerato, acceso desde un módulo de Formación Profesional, prueba de acceso para mayores de 25 años, acceso desde otra titulación universitaria finalizada y acceso desde otra titulación universitaria inconclusa.

- **Dedicación a la formación.** Definida como variable cualitativa *politómica*, establece tres categorías para diferenciar en el alumnado distintos grados de dedicación a la propia formación universitaria. Las alternativas contempladas son tres, y parten de la definición del término *trabajo*, como ocupación retribuida (Real Academia Española, 2014): dedicación exclusiva a la formación, combinación de los estudios universitarios con un trabajo legalmente regulado (relación contractual), y combinación de la formación con un trabajo no regulado (sin contrato).

3.2.4 Posibles fuentes de influencia.

Este cuarto bloque de variables recoge factores que, por su relación con la evaluación del aprendizaje, pueden condicionar la valoración que se realice de la competencia evaluativa docente y las necesidades formativas asociadas. Representan factores con una influencia probable en las valoraciones de la muestra, por lo que se ha practicado un control mínimo sobre las mismas que, al menos, procura su identificación.

Las posibles fuentes de influencia registradas en la **muestra docente** son las tres siguientes:

- **Autopercepción de competencia evaluativa.** Presentada como variable cualitativa *politómica*, propone una escala ordinal para que el profesorado determine el nivel de competencia que se autoatribuye en la evaluación del aprendizaje. Incluye tres alternativas: muy competente, competente y poco competente.
- **Realización de actividades de formación permanente vinculada a la evaluación en los dos últimos años.** Representa una variable cualitativa dicotómica. Detalla si en los dos últimos años se ha llevado a cabo o no alguna actividad formativa vinculada a la evaluación del aprendizaje.
- **Necesidad de formación evaluativa percibida.** Presentada como variable cualitativa dicotómica, divide la muestra entre quienes perciben necesidad de formación permanente para el ejercicio de la evaluación y quienes no sienten tal necesidad.

En la **muestra de estudiantes**, los factores registrados por su posible efecto en las valoraciones vinculadas a la competencia evaluativa son los siguientes:

- **Experiencia previa con la/el docente de referencia.** Esta variable cualitativa dicotómica plantea en una escala nominal si el alumnado ha tenido o no experiencia previa con la/el docente cuya competencia evaluativa se valora. Esta variable se ha incorporado porque buena parte del profesorado trabaja en cursos y asignaturas diferentes a lo largo de la titulación.
- **Calificaciones/resultados parciales recibidos.** Se trata de una variable cualitativa dicotómica. Constata si en el momento de realizar la recogida de información las/os estudiantes cuentan o no con alguna valoración o resultado parcial de su aprendizaje en el módulo/asignatura de referencia.
- **Expectativa de logro en el módulo/ asignatura cursada.** Esta tercera variable hace referencia al resultado académico final que prevé alcanzar la/el estudiante en el momento de realizar la recogida de información. Se presenta como variable cualitativa *politómica* sujeta a tres alternativas de respuesta: preveo superar el módulo/asignatura (expectativa positiva), preveo suspender (expectativa negativa) o tengo dudas al respecto.
- **Percepción general de competencia docente.** Este factor representa una variable cualitativa *politómica* que recoge la percepción general de competencia docente que atribuye el alumnado al profesor/a cuya praxis evaluativa se valora. Propone tres opciones en escala ordinal: poco competente, competente y muy competente.
- **Modalidad de aplicación del cuestionario.** Se trata de una variable independiente sujeta al propio proceso de recogida de datos y responde al efecto potencial de una irregularidad en su desarrollo. Como más tarde será detallado, la planificación inicial preveía la administración directa de los cuestionarios discentes en una sesión de trabajo de 45 minutos. Pensada para evitar la divergencia de interpretaciones entre el alumnado, la sesión planteaba la explicación general de los indicadores y la respuesta simultánea, individual y anónima, por parte del alumnado. Sin embargo, en dos de las recogidas efectuadas, imperativos prácticos obligaron a reducir a 30 minutos el tiempo de trabajo; las/os estudiantes de estos grupos participaron en una modalidad abreviada, donde la respuesta a la parte final del cuestionario se realizó de forma autónoma (sin aclaración previa). En consecuencia, esta variable deja constancia de si la recogida de información practicada en cada caso ha sido abreviada (30 minutos, respuesta autónoma en la parte final) o extensa (45 minutos, aclaración previa en todos los ítems).

3.2.5 Variables extrañas.

La complejidad natural de los fenómenos y contextos educativos los hace especialmente sensibles al influjo de variables extrañas de diversa índole (Kerlinger, 2002). El estudio empírico que está siendo presentado propone la evaluación de una de las competencias clave vinculadas a la docencia. Por proximidad, las influencias reportadas en la literatura para los procesos de evaluación docente, a cargo de estudiantes o del propio profesorado, pueden ofrecer una orientación para anticipar las principales amenazas en este sentido.

La profundización en las fuentes que amenazan el rigor de los procesos de evaluación de la calidad docente excede los propósitos de esta investigación. Sin embargo, la consideración de los trabajos clásicos (Brodie, 1998; Marsh, 1984, 1987) y de algunas de las revisiones publicadas sobre el tema (Cisneros-Cohernour & Stake, 2010; De Juan et al., 2007; B. García et al., 2008; B. García et al., 2012; Richardson, 2005; Williams & Brennan, 2003) deja entrever la incidencia recurrente de ciertos factores. Admitido que no son los únicos, la tabla 62 resume los principales riesgos de imprecisión o sesgo que cabe esperar en la autoevaluación y evaluación de la competencia docente. Aunque en la literatura especializada no existe consenso al explicar la intensidad y la dirección de su efecto, todos ellos representan referencias constantes en el debate sostenido.

Tabla 62

Principales amenazas en los procesos de evaluación de la calidad docente con incidencia potencial en la evaluación o autoevaluación de la competencia evaluativa

AUTOEVALUACIÓN DOCENTE	• Falta de objetividad (tendencia a producir valoraciones por encima de lo constatado externamente por otras fuentes). • Competencia autopercebida en la materia (la tendencia es que a menor competencia autopercebida, menores puntuaciones se registran).
EVALUACIÓN DOCENTE A CARGO DE ESTUDIANTES	• Interés por la asignatura/módulo y tipo de curso (si la asignatura es optativa o de su interés, se producen valoraciones ligeramente más altas que si es obligatoria o no se considera interesante). • Estímulo intelectual y utilidad (el alumnado tiende a valorar mejor las asignaturas que considera intelectualmente desafiantes y útiles y más bajo aquellas otras más fáciles en las que aprende menos). • Motivación (el alumnado tiende a dar puntuaciones más altas cuanto más motivado esté y más sienta que aprende). • Calificaciones esperadas (el alumnado que espera recibir buenas calificaciones tiende a realizar mejores valoraciones). • Dificultad percibida (la tendencia es que cuanto mayor sea el nivel de dificultad percibido, menores sean las valoraciones). • Carga de tareas y calificaciones adjudicadas (la tendencia general indica que a mayores calificaciones y menor cantidad de tareas, mejores son las valoraciones).

EVALUACIÓN DOCENTE A CARGO DE ESTUDIANTES	• Estilo docente estricto o indulgente (la tendencia apunta a una mejor valoración de los estilos de enseñanza indulgente frente a los estrictos). Determinados estudios vinculan este factor con la etnia de la/el docente (se aprecia una mayor predisposición a la indulgencia entre latinas/os). • Condiciones estructurales (condiciones de administración de la prueba, las/os encuestadoras/es y el propio instrumento pueden determinar diferencias sin que se pueda precisar dirección). • Edad (aunque los resultados no sean del todo concluyentes, la edad puede condicionar la credibilidad y adecuación de las valoraciones planteadas; por lo general, se desaconseja su uso con escolares). • Tamaño del grupo (en grupos pequeños, tienden a concentrarse mejores valoraciones). • Rasgos de personalidad del profeso o profesora (aunque los resultados no son del todo concluyentes, parece que determinados rasgos docentes pueden afectar en la tendencia de respuesta). • Personalidad de las/os estudiantes que responden (también sin resultados del todo concluyentes, ciertas investigaciones plantean que lo que condiciona la tendencia de respuesta es la personalidad del alumnado que responde).
EN AMBAS MODALIDADES	• Contexto institucional (se contempla la posible influencia del contexto, entendiendo que en él median la filosofía y propósitos generales, normativa y gestión académica y cultura institucional). • Dispositivos institucionales de evaluación y formación docente (dependiendo de los propósitos desde los que se articule la evaluación y los enlaces que se articulen con la formación). • Disciplina valorada (la naturaleza de la disciplina más o menos técnica, práctica, etc. puede condicionar valoraciones diferentes). • Las características del instrumento (tanto en evaluaciones como autoevaluaciones, este aspecto puede incidir).

Nota: Reelaboración propia a partir de varias contribuciones (Brodie, 1998; García et al., 2008, 2012; Marsh, 1984, 1987).

Sin negar su importancia e incidencia, la presente investigación ha rehusado la consideración de la lista completa de factores por cuestiones de viabilidad. Aquellos en los que se ha estimado una mayor incidencia y posibilidad de constatación han sido presentados en el bloque de variables clasificado como *fuentes de influencia*.

4. Instrumentos para la recogida de la información

De acuerdo con Albert (2009), los cuestionarios representan el nexo entre los objetivos de la investigación y la muestra consultada, por lo que su selección y configuración supone una decisión crítica en la fase de diseño. “Un cuestionario es un

instrumento de recopilación de información compuesto de un conjunto limitado de preguntas mediante el cual el sujeto proporciona información sobre sí mismo y sobre su entorno” (Torrado, 2009, p. 240). Frente a los test, “los términos escala, inventario o cuestionario hacen referencia a instrumentos que pretenden cuantificar características en las que no hay respuestas correctas o incorrectas, como ocurre en la medición de atributos relacionados con las actitudes, opiniones, intereses, disposiciones o personalidad” (Prat & Doval, 2005, p. 45). Tal es el caso de las herramientas empleadas en la presente investigación, centradas en la recogida de percepciones sobre la importancia y el grado de realización de los indicadores recogidos en el modelo de competencia evaluativa docente y las necesidades formativas asociadas.

La presente investigación consulta dos fuentes de información y, en consecuencia, se han diseñado dos versiones diferentes de cuestionario: una para estudiantes y la otra para profesorado. En ambos casos, la estructura de la herramienta incorpora tres bloques diferenciados de cuestiones: los datos de identificación, la valoración del grado de importancia y realización de los indicadores de competencia evaluativa docente cuestionados y las preguntas abiertas finales. El segundo bloque representa la parte principal y deriva directamente del modelo de competencia evaluativa ideado en respuesta a los propósitos teóricos de la investigación (objetivo D; capítulo V). Los indicadores previstos en el mismo pasan a constituir los ítems del cuestionario y requieren una doble respuesta para valorar su grado de relevancia y cumplimiento práctico. Las categorías o alternativas de respuesta referidas a la importancia son comunes en todos ellos (*irrelevante, poco relevante, relevancia moderada, relevante y muy relevante*), pero con la realización no sucede igual. En este caso se ha recurrido a la configuración de la rúbrica o rejilla de valoración que representa de forma descriptiva el significado de cinco niveles de realización en cada uno de los indicadores evaluados (*inadecuado, poco adecuado, adecuado, muy adecuado y excelente*). En ambos casos la categoría central (3, *relevancia moderada y cumplimiento adecuado*) se corresponde con la media teórica.

Las rúbricas o matrices de valoración representan guías de puntuación empleadas para la evaluación del desempeño; detallan las características específicas de un producto, proyecto o tarea en varios niveles de rendimiento, con el fin de clarificar expectativas, realizar valoraciones y facilitar la retroalimentación (Andrade, 2000, 2005; Blanco, 2008b). En otras palabras, “la rúbrica proporciona una descripción de los distintos niveles de desempeño cuantitativo para una tarea de rendimiento y describe lo que el dominio de dicha tarea (y sus diferentes grados de dominio) debería representar” (Hafner & Hafner, 2003, p. 1510). Su empleo en los últimos años se ha visto notablemente incrementado por la expansión de los enfoques evaluativos basados en el desempeño que propicia la

aproximación a las competencias (De la Orden, 2011a, 2011b). Entre otros beneficios, contribuyen a centrar la atención y a unificar significados sobre los aspectos a valorar y ese propósito responde su integración en los cuestionarios diseñados.

Estudiantes y profesorado sostienen una posición y vivencia desigual con respecto a la evaluación del aprendizaje. En consecuencia, aunque el punto de partida sea común en ambas herramientas, existen diferencias en la forma y en las cuestiones consultadas en cada caso.

El *cuestionario de autoevaluación ideado para recabar la percepción del profesorado* sobre su propia competencia evaluativa y las necesidades formativas relacionadas se presenta en el **ANEXO 14**. Su estructura se ajusta exactamente al patrón introducido anteriormente y ha sido respondido de forma autónoma por parte de las/os participantes (ver tabla 63):

Tabla 63
Estructuración del cuestionario de autoevaluación de la competencia evaluativa docente

BLOQUE	DESCRIPCIÓN
INTRODUCCIÓN	Descripción de los propósitos centrales de la investigación y de las instrucciones de cumplimentación.
IDENTIFICACIÓN	Recogida de 10 datos vinculados a las variables demográficas, institucionales, académicas y profesionales. Hay cuatro datos más que se proporcionan directamente según lo acordado en el proceso de muestreo para evitar confusión (nombre del módulo/asignatura de referencia, titulación, curso, módulo y materia).
INDICADORES	Valoración del grado de importancia y de la realización autopercibida en los 19 indicadores de competencia evaluativa previstos en el modelo teórico y en la rúbrica de referencia (ANEXO 05), con dos adecuaciones menores: - Se incorpora en una fila la valoración de la importancia atribuida a cada indicador (en una versión reducida para no extender en exceso el formulario y evitar la dispersión). - El contenido de la rúbrica se redacta en primera persona para reforzar el sentido de autoevaluación y los calificativos sobre el grado de realización se sustituyen por números.
PREGUNTAS FINALES	- Dos preguntas abiertas: puntos fuertes y áreas de mejora vinculadas a la evaluación del aprendizaje y necesidad de formación permanente (si se siente o no tal necesidad y una descripción de la demanda en su caso). - Una pregunta semi-abierta: descripción de la formación evaluativa desarrollada en los dos últimos años, si procede; para orientar mínimamente las respuestas y favorecer su posterior categorización, contempla una serie de preguntas cerradas sobre dicha formación (duración, obligatoriedad u optatividad y satisfacción con la misma).

El cuestionario destinado a recoger la percepción del alumnado sobre la competencia evaluativa docente se expone en el **ANEXO 15**. Como se desarrolla más adelante, su administración ha tenido lugar en una sesión dirigida para facilitar entre las/os estudiantes la interpretación unívoca de los indicadores valorados. La estructura del instrumento también se ordena en tres bloques, pero con ciertas modificaciones para adecuar el contenido a las características de la fuente (ver tabla 64):

Tabla 64
Estructuración del cuestionario de evaluación de la competencia evaluativa docente para estudiantes

BLOQUE	DESCRIPCIÓN
INTRODUCCIÓN	Descripción de los propósitos centrales de la investigación y de las instrucciones de cumplimentación.
IDENTIFICACIÓN	Recogida de nueve datos vinculados a las variables demográficas, institucionales, académicas y profesionales. También aquí el nombre del módulo/asignatura de referencia, la titulación y el curso aparecen precisados.
INDICADORES	Determinados aspectos de la planificación de los procesos evaluativos o la justificación de algunas decisiones vinculadas a su implementación o revisión tienen lugar en un espacio reservado de la docencia, de modo que no son evaluables por el alumnado (ANEXO 16). La supresión de los indicadores ajenos al alumnado y la adecuación de los que sólo pueden ser valorados parcialmente han dado lugar a un banco de 22 indicadores de la competencia evaluativa docente en esta versión^a. Su contraste se realiza en dos formatos diferentes: • <i>Rúbrica de valoración para estudiantes</i>. En 11 indicadores se solicitan valoraciones en un formato simplificado de rúbrica. Redactada desde la perspectiva discente, ofrece descripciones más reducidas y recurre a un lenguaje menos técnico. • <i>Ítems tipo importancia-realización de escala fija</i>. Aquellos indicadores en los que no es posible la valoración de todos los niveles de realización previstos en la rúbrica original (los que componen las dimensiones cuarta, quinta y octava, según lo precisado en el ANEXO 16), han sido convertidos en 11 ítems convencionales de tipo importancia-realización^b.
PREGUNTAS FINALES	El tercer bloque del cuestionario lo conforma una única pregunta abierta destinada a captar los puntos fuertes y áreas de mejora que advierte el alumnado en la evaluación de su aprendizaje.

^a Para tomar conciencia de la magnitud del cambio, el **ANEXO 17** detalla el reflejo de los indicadores de competencia evaluativa previstos en el modelo teórico en ambos cuestionarios.

^b Los indicadores incluidos en esta sección no ofrecen información directamente comparable con la obtenida en la muestra docente, pero proporcionan información de contraste útil para la elaboración de los informes de las/os participantes.

Presentadas de forma general ambas versiones del cuestionario, los próximos apartados exponen las fases recorridas en su proceso de creación y sus principales características psicométricas (validez y fiabilidad). El punto final lo pone una reflexión

realista sobre sus límites y posibilidades para contribuir a los fines de la presente investigación.

4.1 Proceso de creación de las herramientas.

La definición del modelo de competencia evaluativa de la que parten ambas versiones del cuestionario es producto de un proceso de revisión y contraste que ha sido expuesto en el capítulo quinto del marco teórico. La creación de las herramientas de recogida de información presentadas a ambos sectores de la muestra debe considerarse como lógica continuación de dicha secuencia:

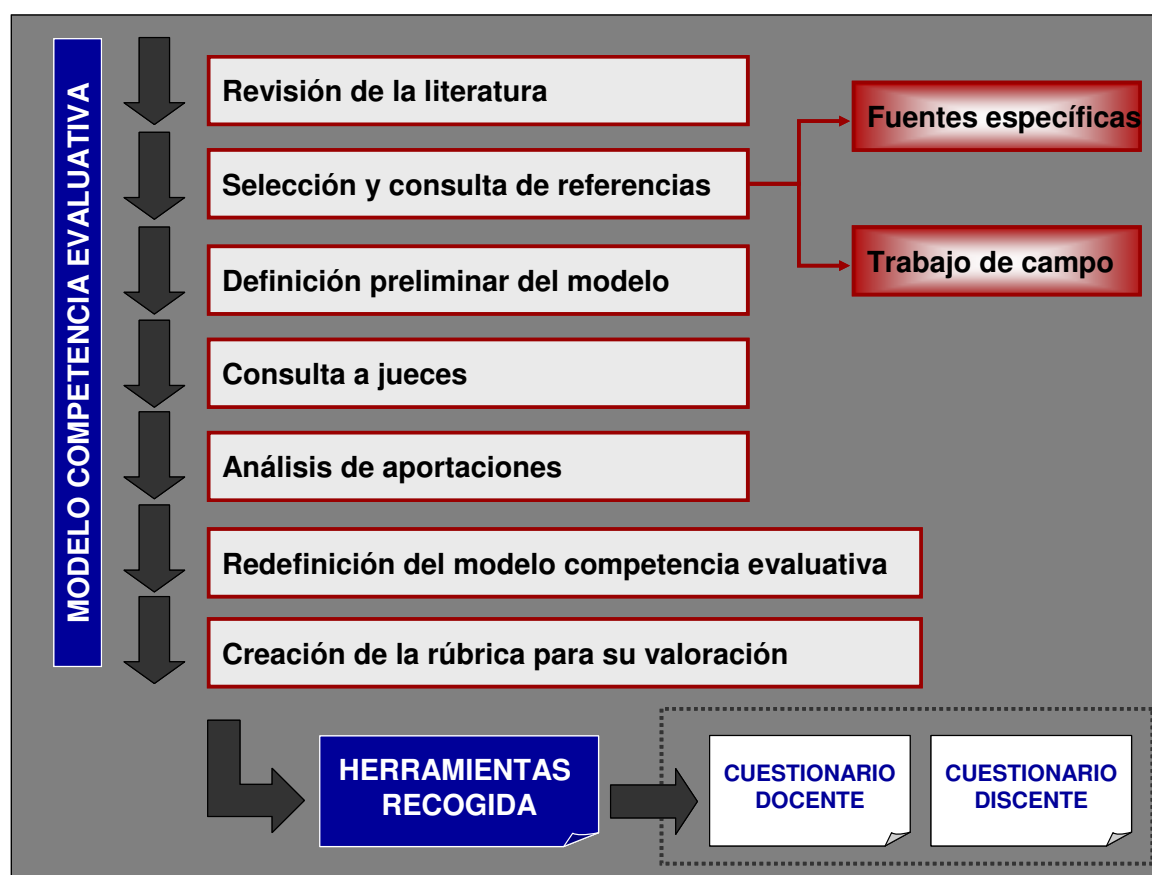


Figura 29. La creación de las herramientas de recogida de información en el marco general de la investigación.

Como refleja la figura 29, ambas versiones del cuestionario parten de la rúbrica que concreta indicadores y descriptores para la competencia evaluativa del profesorado. Desde ese marco común de referencia, el diseño de los cuestionarios se ha llevado a cabo

siguiendo los mismos pasos. Como detalla la tabla 65, su desarrollo enlaza las aportaciones procedentes del marco teórico con distintos momentos de contraste empírico:

Tabla 65

Proceso de creación del cuestionario en la versión docente

FASE	DESCRIPCIÓN
1. SELECCIÓN DE PREGUNTAS	• VERSIÓN DOCENTE. Adecuación de la rúbrica de referencia a la perspectiva docente (primera persona e incorporación de la relevancia) y selección de las preguntas adecuadas para recoger información sobre el resto de variables dependientes y los datos de identificación. • VERSIÓN DISCENTE. Adecuación de la rúbrica discente (redacción simplificada y adecuación del lenguaje) y formulación de ítems tipo importancia-realización en escala de respuesta. También la selección de preguntas necesarias para recoger información sobre el resto de variables (datos de identificación y valoración global de las fortalezas y debilidades asociadas a la evaluación).
2. MAQUETACIÓN DE LA PRIMERA VERSIÓN	Partiendo de la selección anterior, en la segunda fase se ha configurado una primera versión de ambos cuestionarios.
3. APLICACIÓN PILOTO	Para valorar la aplicabilidad, claridad y adecuación de la primera versión de los cuestionarios, se ha llevado a cabo una prueba piloto para cada una de las versiones. • VERSIÓN DOCENTE. La aplicación piloto se ha realizado contando con la colaboración de 16 profesoras/es vinculados a la formación básica de maestras/os. Para evitar sesgos o posibles interferencias con la investigación, se ha recurrido a profesorado no susceptible de ser incluido en la muestra (docentes de otros centros y comunidades autónomas o sin ocupación docente actual). El ANEXO 18 presenta el listado de participantes, así como las principales conclusiones obtenidas en la aplicación. • VERSIÓN DISCENTE. La prueba se ha llevado a cabo contando con la colaboración de un grupo de 29 estudiantes del primer curso de Educación Primaria. La elección de estudiantes de primer año ha sido adoptada de forma deliberada, presuponiendo que se trata del nivel donde mayores dificultades puede plantear la herramienta desde la perspectiva terminológica. Aplicando las recomendaciones de Prat y Doval (2005), las/os integrantes del mismo han sido automáticamente excluidos de la configuración de la muestra. El ANEXO 19 resume las principales conclusiones extraídas de esta experiencia piloto. En este caso, no se detalla la identidad de las/os participantes ya que respondieron de forma anónima, replicando la aplicación prevista.
4. REVISIÓN: SEGUNDA VERSIÓN	Analizadas las aportaciones desprendidas de la fase piloto, se ha configurado una segunda versión de ambos cuestionarios.
5. APLICACIÓN DEL CUESTIONARIO: PRIMERA FASE	Los cuestionarios creados en la fase anterior son los empleados para recopilar información de la muestra docente y discente, de modo que su primera aplicación tiene lugar entre abril y mayo de 2012.

FASE	DESCRIPCIÓN
6. ANÁLISIS TENTATIVO	Partiendo de los datos recopilados en la primera fase de la recogida, se ha realizado un análisis tentativo de la validez y la fiabilidad. El objetivo del mismo ha sido descartar problemas psicométricos que pudieran poner en entredicho los resultados de la consulta. Confirmada esta premisa, se asume la provisionalidad de los datos hasta finalizar la recogida.
7. SEGUNDA FASE DE APLICACIÓN	La segunda fase de la recogida de datos ha tenido lugar entre noviembre y diciembre de 2012 empleando las mismas herramientas.
8. ANÁLISIS DE RASGOS PSICOMÉTRICOS	Completada la recogida de información, se ha llevado a cabo el estudio definitivo de fiabilidad y validez de las respuestas obtenidas con ambos cuestionarios. Las particularidades de dicho estudio y sus conclusiones se presentan en los dos próximos apartados.
9. REDACCIÓN DE LA RÚBRICA DEFINITIVA	Analizadas las limitaciones detectadas en esta primera aproximación al análisis de la competencia evaluativa docente, el ANEXO 20 deja planteadas una serie de modificaciones sobre la rúbrica de referencia para mejorar la estructura y consistencia interna de las respuestas en próximas investigaciones.

4.2 Validez.

En términos generales, la validez es la capacidad de la escala o cuestionario para medir aquello que dice medir (Belmonte, 2002; G. Pérez, 1994; Prat & Doval, 2005), dando lugar a hallazgos auténticos, es decir, isomórficos a la realidad a la que hacen referencia y confiables (Denzin & Lincoln, 2012). Esta representación general del término parece asumida y aceptada de forma general, pero su profundización da lugar a una extensa variedad de acepciones e interpretaciones.

Ampliamente reconocidos en la investigación vinculada a las ciencias sociales, este trabajo adopta la última edición de los estándares de la AERA (American Educational Research Association), la APA (American Psychological Association), y la NCME (National Council on Measurement in Education) como marco de referencia para su justificación psicométrica. En ellos, la validez se define como el grado en que la evidencia empírica y la literatura apoyan la interpretación de los resultados del cuestionario para su uso previsto (AERA, APA, & NCME, 2014).

Profundizando en las posibles interpretaciones, la AERA, APA y NCME (2014) reconocen cinco grandes modalidades de validez generalmente aplicables a la investigación

educativa. En atención a la viabilidad del proyecto y sus propósitos prioritarios, la investigación presentada responde particularmente a tres de ellas:

- **Validez basada en el contenido.** Hace referencia a la capacidad del instrumento de recogida para representar de forma suficiente las distintas facetas que dan lugar al constructo evaluado, en nuestro caso, la competencia evaluativa docente. Está sujeta a dos condiciones necesarias: la relevancia, que hace referencia a que los ítems estén directamente relacionados con el atributo a medir, y la representatividad, que hace referencia a la capacidad de la muestra de ítems para recabar información sobre las distintas facetas implicadas en el constructo. El incumplimiento de la primera condición daría lugar a sesgo por irrelevancia de constructo, mientras que la insatisfacción de la segunda ocasionaría sesgo por infra representación (Goodwin & Leech, 2003; Prat & Doval, 2005; G. Prieto & Delgado, 2010).

Este tipo de validez ha sido preservada durante la fase de configuración del modelo de competencia evaluativa docente, recurriendo a fuentes variadas y relevantes para respaldar su definición. Entre ellas ha sido incluida la consulta a expertas/os, una de las estrategias más frecuentes para favorecer la validez de contenido (Prat & Doval, 2005). Antes de proceder a la recogida, también se han realizado aplicaciones piloto con ambos sectores de la muestra para procurar la representación de toda la casuística posible y detectar posibles interferencias.

- **Validez basada en la estructuración interna.** Se produce cuando la organización real de los ítems en los resultados de la aplicación se ajustan a lo previsto en la definición del atributo que pretende medirse. Es decir, “el análisis de la estructura interna persigue verificar empíricamente si los ítems se ajustan a la dimensionalidad prevista por el constructor de la prueba” (G. Prieto & Delgado, 2010, p. 72).

En la presente investigación este tipo de validez ha sido delimitada por medio de un análisis factorial exploratorio realizado para las subescalas de importancia de ambos cuestionarios. Dicho análisis ha dado lugar a la identificación de seis factores en cada caso, estrechamente relacionados con las ocho dimensiones previstas inicialmente en el modelo teórico de referencia. El procedimiento de análisis, así como la factorización resultante, se presentan en el siguiente apartado.

- **Validez basada en las consecuencias.** Este tipo de validez está asociada a las consecuencias que deja tras de sí la aplicación del cuestionario y la estimación de resultados. Para favorecerla, la estrategia aplicada en la presente investigación, ha sido fundamentalmente la información previa y sensibilización sobre el carácter

formativo de la experiencia en ambos sectores de muestra. Dicho carácter, unido a la voluntariedad y al compromiso de anonimato, aleja los cuestionarios empleados de lo que la literatura psicométrica denomina como usos de alto riesgo (G. Prieto & Delgado, 2010).

4.2.1 Análisis factorial exploratorio de los cuestionarios docente y discente para la determinación de la validez de estructura.

“Con el término análisis factorial hacemos referencia a un conjunto de técnicas estadísticas que tienen como objeto proveer las bases empíricas para reducir un gran número de variables a un número menor de variables hipotéticas denominadas factores” (Mateo, 2009, p. 217). Su objetivo es simplificar las numerosas y complejas relaciones que se puedan encontrar en un conjunto de variables cuantitativas observadas, recurriendo a la identificación de factores (De Vicente & Manera, 2005). Este análisis puede ser confirmatorio o exploratorio en función de que el investigador disponga o no de una hipótesis sobre el número de factores comunes a priori. El primero realiza comprobaciones de hipótesis sobre las cargas factoriales, mientras que el segundo consiste en una exploración de los resultados obtenidos para la identificación de factores.

Tanto el cuestionario de autoevaluación docente como el dirigido al alumnado han sido contruidos sobre la base de un modelo teórico de ocho dimensiones, aquel que desglosa la competencia evaluativa del profesorado de referencia. No obstante, se ha llevado a cabo un análisis exploratorio para identificar la estructura de factores que revelan las respuestas recabadas en ambas muestras. Como señalan G. Prieto y Delgado (2010), no hacerlo daría lugar a interpretaciones desajustadas y erróneas de los resultados asociados al objeto de estudio.

En ambos cuestionarios los indicadores de la competencia evaluativa son sometidos a una doble valoración, dando lugar a dos subescalas: la referida a la importancia y la centrada en la realización percibida. Técnicamente, el análisis factorial reúne las variables (ítems) con correlaciones moderadas o altas entre sí para identificar los factores (Mateo, 2009). Los propósitos de la investigación llevan a analizar la competencia evaluativa de diferentes profesoras/es en módulos o asignaturas diversas dentro del plan de estudios. Las/os participantes no realizan la valoración de una misma situación, por lo que en la subescala de realización son previsibles grandes discrepancias dependiendo del contexto en el que se haya focalizado el análisis (sistema evaluativo de un/a representante docente, en un módulo/asignatura particular). En consecuencia, la exploración de la estructuración

interna de las respuestas se ha llevado a cabo sobre los ítems que conforman la subescala de importancia, en los que se presupone una respuesta más general o estable.

Las soluciones factoriales halladas para cada caso se resumen en los dos próximos subapartados, empezando por la docente y siguiendo con la del alumnado.

4.2.1.1 *Análisis factorial exploratorio de la subescala de importancia del cuestionario dirigido al profesorado.*

En el cuestionario docente la subescala de importancia la componen 19 ítems. Hallada la matriz de correlaciones, el método de extracción por análisis de componentes principales ha identificado seis factores que, en suma, explican el 77,43% de la varianza total (ver tabla 66). Dicho método analiza la varianza común y no común entre variables y es el más recomendable al analizar una cantidad considerable de elementos (en torno a 20, según Nunnally, 1978); el criterio aplicado para la retención de los componentes ha sido la obtención de autovalores superiores a 1, por ser la alternativa más conservadora y la que mayor varianza refleja (De Vicente & Manera, 2005; Morales, 2013):

Tabla 66
Extracción de componentes principales y varianza total explicada en la subescala de importancia en docentes

	VARIANZA TOTAL EXPLICADA								
	Autovalores iniciales			Suma de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	TOTAL	% de la varianza	% acumul.	TOTAL	% de la varianza	% acumul.	TOTAL	% de la varianza	% acumul.
1	6,283	33,069	33,069	6,283	33,069	33,069	2,946	15,505	15,505
2	2,505	13,184	46,252	2,505	13,184	46,252	2,917	15,351	30,856
3	2,021	10,637	56,890	2,021	10,637	56,890	2,817	14,825	45,681
4	1,692	8,908	65,797	1,692	8,908	65,797	2,476	13,030	58,711
5	1,193	6,278	72,075	1,193	6,278	72,075	2,187	11,512	70,223
6	1,018	5,357	77,432	1,018	5,357	77,432	1,370	7,208	77,432

Para valorar la pertinencia de la solución, se ha procedido a rotar los factores, por medio del sistema ortogonal Varimax, cuyo objetivo es minimizar el número de variables con cargas altas en un factor (De Vicente & Manera, 2005). La tabla 67 recoge el resultado:

Tabla 67

Matriz de componentes rotados en la subescala de importancia en docentes

Ítem	Matriz de componentes rotados (Varimax con Kaiser)					
	COMPONENTE					
	1	2	3	4	5	6
C.1.1.	,887	,099	,226	,008	-,075	,157
C.1.2.	,780	,283	-,131	,058	-,015	,130
C.1.3.	,609	,131	,198	,153	,200	,224
C.1.4.	,686	-,011	,084	,251	,560	-,062
C.5.3.	,093	,817	-,072	,104	,227	,260
C.6.1.	,139	,886	,222	,089	-,063	,058
C.6.2.	,098	,729	,462	,176	,138	-,124
P.6.3.	,311	,652	,266	,157	-,254	-,285
P.8.1.	,127	,170	,827	-,059	,347	-,007
P.8.2.	,126	,280	,731	,097	,277	,189
P.8.3.	,018	,082	,870	,216	,032	,056
C.3.1.	,120	,041	,288	,764	,091	,023
C.3.2.	-,405	,013	,209	,729	,094	,278
P.4.2.	,266	,235	-,123	,730	,088	,081
C.7.1.	,355	,278	-,022	,595	,298	,048
P.5.1.	,102	,055	,306	,099	,874	-,021
P.5.2.	-,063	,007	,354	,423	,684	,113
C.2.1.	,292	,381	,007	,046	,362	,582
P.4.1.	,261	-,073	,165	,244	-,123	,794

Para valorar la adecuación de la estructura interna resultante no existe un único criterio, pero la consideración conjunta de orientaciones teóricas y psicométricas nos lleva a determinar que la solución presentada es razonable para responder a los propósitos de la investigación.

Desde el punto de vista psicométrico, la adecuación de la solución representada en la tabla 67 ha sido valorada en los siguientes términos:

- Interpretadas como coeficientes de correlación, las *cargas de las variables* que integran cada factor se consideran muy aceptables y no ofrecen indicios de irrelevancia. Según Morales (2013), las cargas por encima de .50 denotan una clara relevancia práctica del ítem dentro del factor, estimación que también comparten Costello y Osborne (2005). En una revisión amplia del tema, Beavers et al. (2013) presentan estimaciones mucho más exigentes. Sin embargo, considerando que los propósitos de la investigación no contemplan la generalización de resultados, se ha adoptado como referencia la primera estimación, que se confirma cubierta en todos

los casos. Aplicadas las orientaciones que proporcionan Bisquerra (1987) o Morales (2008a) para determinar la magnitud de las correlaciones (ver tabla 68), los 19 ítems presentan un peso moderado, alto o muy alto dentro de los factores.

Tabla 68

Orientaciones para interpretar el tamaño de la relación según Bisquerra (1987) y Morales (2008a)

Coefficiente de correlación	Mayor a .80	Muy alta	ASOCIACIÓN
	Entre .60 y .79	Alta	
	Entre .40 y .59	Moderada	
	Entre .20 y .39	Baja	
	Inferior a .20	Muy baja	

- En cuanto al *número de componentes por factor*, Costello y Osborne (2005) o Morales (2013) determinan que ningún factor debería estar conformado por menos de tres ítems. Este requisito no se cumple en los dos últimos casos por la propia configuración de la rúbrica de referencia, que incorpora dos dimensiones con un único indicador⁸¹.
- Morales (2013) define como factor sólido aquel conformado por cinco ítems con cargas superiores a .50; aunque el número de elementos no llegue a cinco en ningún caso, los cuatro factores restantes contienen tres y cuatro variables (indicadores o ítems) con saturaciones muy por encima de ese nivel.
- Atendiendo a la *claridad general de la estructura*, se aprecia que las diferentes variables aparecen saturadas por un único factor como cabe esperar de una solución adecuada, aceptable y sin ambigüedad (Prat & Doval, 2005). Una estructura clara es aquella en la que los ítems que definen un factor tienen pesos por encima de .50 en el mismo y menores en el resto (Morales, 2013) y esta orientación se satisface sin excepción.

Desde el punto de vista teórico, la adecuación de la solución presentada también ha sido valorada en función de su *significación conceptual* (Morales, 2013). El análisis factorial sólo representa una solución matemática entre otras muchas posibles, por lo que su elección no debe desligarse de la naturaleza de la investigación (Mateo, 2009). Aunque la

⁸¹ Como ya ha sido justificado, la configuración de una rúbrica de evaluación con cinco descriptores por indicador ha llevado a limitar la extensión del cuestionario en lo posible, para reducir los riesgos asociados a la fatiga. Dicha reducción supone asumir dificultades psicométricas que ya han sido avanzadas, y sobre las que volveremos a insistir en la presentación de los factores.

coincidencia no sea plena, apreciamos en este sentido que los seis factores delimitados sintetizan y representan adecuadamente las ocho dimensiones previstas en la propuesta teórica de partida (ver figura 30):

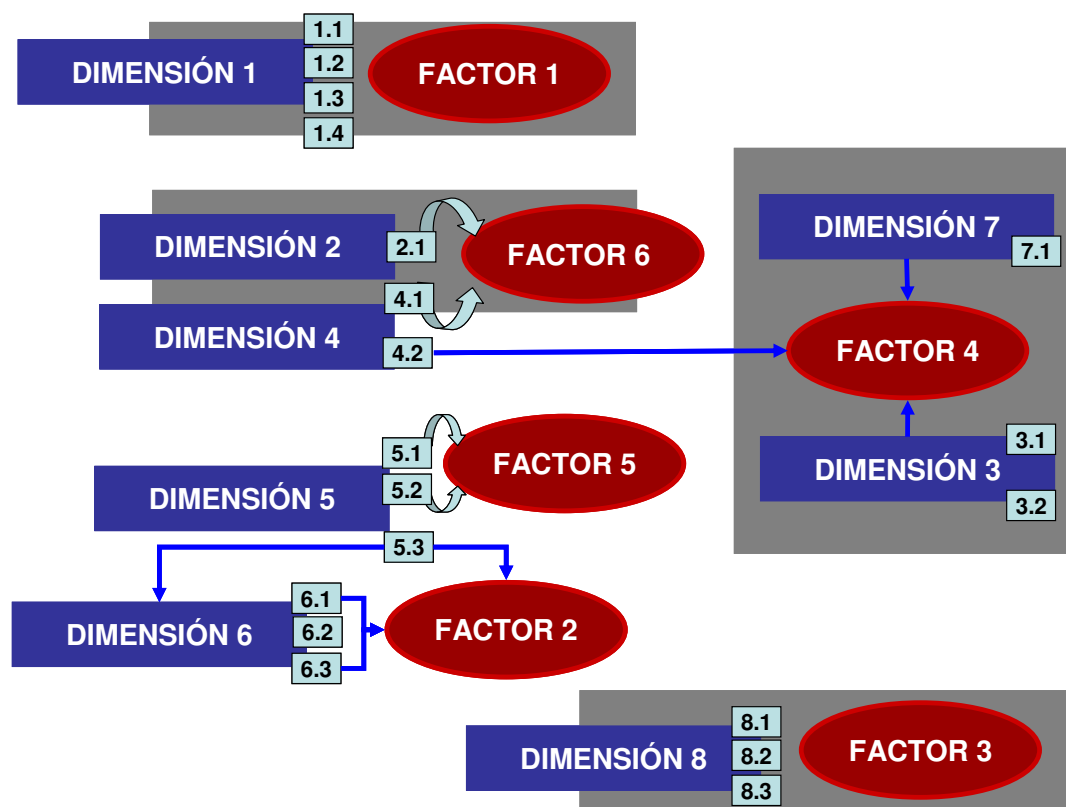


Figura 30. Estructura interna de la respuesta docente y relación con el modelo teórico de referencia.

Esta significación teórica y conceptual se evidencia al explorar la descripción de los factores resultantes:

Factor 1. Planificación estratégica de la evaluación.

El primer factor hace referencia a la planificación estratégica del aprendizaje y se corresponde plenamente con la primera dimensión presentada para el modelo teórico (ANEXO 05). Los cuatro ítems o indicadores que conforman el factor son comunes para alumnado y profesorado⁸² y muestran correlaciones muy altas (C.1.1.) y moderadas (C.1.2., C.1.3., C.1.4.) según las orientaciones proporcionadas (Bisquerra, 1987; Morales, 2008a).

⁸² A ello responde que su codificación comience con un C (Común), de acuerdo a lo establecido en el ANEXO 17.

Recopilados en la tabla 69 según su peso o carga en el factor, cada uno de ellos centra la atención en un aspecto de la planificación evaluativa: la continuidad de la evaluación (C.1.1.), su alineación con objetivos y metodología (C.1.2.), la combinación de las funciones sumativa y formativa (C.1.3.) y la elección de herramientas evaluativas adecuadas para evidenciar el aprendizaje pretendido (C.1.4.).

Tabla 69

Componentes del Factor 1 en la versión docente del cuestionario (Planificación estratégica de la evaluación) y sus cargas factoriales

Ítem	ENUNCIADO	Carga
C.1.1.	<i>Articulo un sistema de evaluación continua que, sensible al proceso de construcción que efectúan las/os estudiantes en su aprendizaje, recabe evidencias en las distintas fases del mismo.</i>	.887
C.1.2.	<i>Planteo una propuesta evaluativa alineada con los objetivos formativos y las estrategias metodológicas previstas para su consecución.</i>	.780
C.1.4.	<i>Preveo el uso de instrumentos y procedimientos de evaluación pertinentes, variados y adecuados al contexto evaluativo específico para obtener evidencias sobre el aprendizaje de las competencias.</i>	.686
C.1.3.	<i>Diseño un sistema evaluativo que combine sus funciones pedagógicas y acreditativas para orientar estratégicamente a las/os estudiantes en el desarrollo de las competencias previstas en el currículo.</i>	.609

Factor 2. Regulación del aprendizaje y la enseñanza por medio de la evaluación.

Este segundo factor hace referencia a la regulación de los procesos de aprendizaje y enseñanza a partir de los resultados puestos de manifiesto en la evaluación del aprendizaje discente. Incluye elementos relacionados con la comunicación de resultados (C.5.3.) y los procesos de feedback (C.6.1.) y proacción o mejora que tienen lugar a partir de la misma, ya sea sobre el aprendizaje discente (C.6.2.) o sobre la enseñanza (P.6.3.).

El factor está compuesto por cuatro ítems, tres comunes en ambas herramientas y uno más que sólo se incluye en el cuestionario docente⁸³. Dichos componentes fusionan la dimensión sexta del modelo teórico completa (Regulación del aprendizaje y la enseñanza) con uno de los indicadores integrados en la dimensión quinta, aquel que se refiere a la comunicación de resultados (ver figura 30). Esta integración adopta pleno sentido desde la perspectiva conceptual; de hecho, brinda respuesta a una de las dudas que se planteó en la configuración del modelo teórico de referencia, donde se valoró que la comunicación de resultados podría estar ligada al análisis e interpretación de las evidencias de aprendizaje

⁸³ Su código aparece precedido de una P (Profesorado) según lo definido en el **ANEXO 17**.

(quinta dimensión del modelo) o al feedback y la proacción, como sugieren las respuestas del profesorado de la muestra.

Las correlaciones de los componentes con el factor son muy altas en los ítems que hacen referencia a la comunicación de resultados y al feedback y altas en los que apelan a las propuestas de mejora para el aprendizaje o la enseñanza (ver tabla 70).

Tabla 70

Componentes del Factor 2 en la versión docente del cuestionario (Regulación del aprendizaje y la enseñanza por medio de la evaluación) y sus cargas factoriales

Ítem	ENUNCIADO	Carga
C.6.1.	<i>Ofrece a las/os estudiantes feedback a partir de las actividades evaluativas desarrolladas.</i>	.886
C.5.3.	<i>Presenta los resultados de forma clara, concisa y bien fundamentada, por medio de formatos que respetan la confidencialidad de las/os estudiantes.</i>	.819
C.6.2.	<i>Especifica pautas claras de proacción para orientar a las/os estudiantes en el desarrollo de las competencias previstas.</i>	.729
P.6.3.	<i>Introduce ajustes para mejorar la enseñanza en función de los resultados de aprendizaje de las/os estudiantes.</i>	.652

Factor 3. Metaevaluación.

En correspondencia con la dimensión octava prevista en el modelo teórico para la sistematización de la competencia evaluativa, este factor centra la atención en la metaevaluación. Incorpora tres ítems o indicadores en referencia a la revisión de los principios éticos (P. 8.3.), pedagógicos (P.8.1.) y técnicos (P. 8.2.) en la evaluación. Representados en la tabla 71, sus correlaciones son altas o muy altas.

Tabla 71

Componentes del Factor 3 en la versión docente del cuestionario (Metaevaluación) y sus cargas factoriales

Ítem	ENUNCIADO	Carga
P.8.3.	<i>Revisa el cumplimiento de los principios éticos de justicia, privacidad, transparencia y beneficencia en el sistema evaluativo.</i>	.870
P.8.1.	<i>Revisa el cumplimiento de los principios de utilidad y consecuencialidad en el sistema evaluativo experimentado.</i>	.827
P.8.2.	<i>Revisa el cumplimiento de los principios de viabilidad, sistematicidad validez y fiabilidad en el sistema evaluativo.</i>	.731

Factor 4. Ajustes de la evaluación y participación discente.

El cuarto factor detectado gira en torno al lugar que ocupa el alumnado en la evaluación e integra en sí dos focos de atención. Por un lado, incluye la dimensión que trata la adecuación de la evaluación a las necesidades discentes (tercera en la propuesta teórica de partida, formada por los indicadores C.3.1. y C.3.2.) y la documentación de los ajustes resultantes (indicador P. 4.2., inicialmente asociado a la dimensión cuarta). Por la otra, incluye también la participación de las/os estudiantes en la evaluación, planteada en la séptima dimensión del modelo teórico (figura 30).

De los tres indicadores que conforman el modelo, tres son comunes en ambos cuestionarios y el cuarto es específico para profesorado (P.4.2.). Los ítems relativos a la adecuación de las condiciones de la evaluación y su consiguiente documentación (C.3.1., C.3.2. y P.4.2.) presentan correlaciones altas. La participación del alumnado en la evaluación correlaciona en menor medida, pero por encima de .50 que es el mínimo adoptado para considerarlo relevante (ver tabla 72).

Tabla 72

Componentes del Factor 4 en la versión docente del cuestionario (Ajustes de la evaluación y participación discente) y sus cargas factoriales

Ítem	ENUNCIADO	Carga
C.3.1.	<i>Determina el nivel de partida de las/os estudiantes con respecto al aprendizaje a efectuar, identificando aspectos que puedan condicionar de forma determinante la evolución y desarrollo del proceso.</i>	.764
P.4.2.	<i>Documenta de forma justificada los ajustes y modificaciones introducidos en el desarrollo del proceso evaluativo en respuesta a los imperativos prácticos e imprevistos.</i>	.730
C.3.2.	<i>Introduce ajustes y medidas para adecuar las condiciones de la evaluación a la diversidad del alumnado garantizando la transparencia del proceso.</i>	.729
C.7.1.	<i>Desarrolla actividades que provocan la participación activa de las/os estudiantes en la evaluación.</i>	.595

El último ítem procede de la dimensión séptima, una de las conformadas por un único indicador para evitar la prolongación de los cuestionarios y los consiguientes riesgos de desatención y fatiga. De ahí que se vincule a indicadores de otras dimensiones dentro de este factor. La configuración resultante sugiere que el profesorado asocia las actuaciones evaluativas que guardan relación directa con el alumnado dentro de un mismo bloque: la consideración de su situación de partida, la introducción y documentación de los ajustes pertinentes, así como su intervención en actividades de autoevaluación, coevaluación y evaluación dialogada, en las distintas fases del proceso evaluativo y con efecto pedagógico y/o acreditativo.

Factor 5. Adecuación y precisión

El quinto factor hace referencia a los rasgos psicométricos de la evaluación del aprendizaje, centrando la atención en la validez y fiabilidad que procura en la misma el profesorado.

El factor lo conforman dos ítems o indicadores específicos para docentes. Presentados en la tabla 73, cada uno de ellos plantea la articulación de medidas para garantizar el cumplimiento de uno de los requisitos: la adecuación o validez del proceso de análisis e interpretación de evidencias (P.5.1.) y su precisión o fiabilidad (P.5.2.). En este caso no se satisface el requisito del número mínimo de componentes; aun así, las cargas de los dos indicadores que incorpora y su alta significación conceptual nos han llevado a considerar compensada tal limitación.

Tabla 73

Componentes del Factor 5 en la versión docente del cuestionario (Adecuación y precisión) y sus cargas factoriales

Ítem	ENUNCIADO	Carga
P.5.1.	<i>Articula deliberadamente medidas y mecanismos para supervisar la validez en el proceso de análisis e interpretación de evidencias.</i>	.874
P.5.2.	<i>Articula deliberadamente medidas y mecanismos para supervisar la precisión en el proceso de análisis e interpretación de evidencias.</i>	.684

Factor 6. Puesta en práctica de la evaluación.

El último factor identificado hace referencia a la implementación o puesta en práctica del sistema de evaluación planificado. Integra dos indicadores vinculados a esta fase del proceso evaluativo (ver tabla 74): el cumplimiento coherente del plan previsto (P.4.1.) y el desarrollo de actividades para la interiorización de los criterios de base por parte del alumnado en su desarrollo (C.2.1.).

Se trata del factor que menos porcentaje de varianza explica y está conformado por dos ítems, incumpliendo las orientaciones sobre el número de componentes necesarios para garantizar su solidez (Costello & Osborne, 2005; Morales, 2013). También integra el

ítem con menor correlación de todos los factorizados, procedente de la segunda dimensión del modelo teórico, conformada por un único indicador o elemento⁸⁴.

Tabla 74

Componentes del Factor 6 en la versión docente del cuestionario (Puesta en práctica de la evaluación) y sus cargas factoriales

Ítem	ENUNCIADO	Carga
P.4.1.	<i>Pone en práctica el sistema evaluativo de acuerdo a las decisiones adoptadas en la fase de planificación, atendiendo a las principales condiciones que median en su aplicación.</i>	.794
C.2.1.	<i>Desarrolla actividades específicas para facilitar la interiorización y apropiación de los criterios de evaluación por parte de las/os estudiantes a lo largo del proceso formativo.</i>	.582

4.2.1.2 **Análisis factorial exploratorio de la subescala de importancia del cuestionario dirigido al alumnado.**

El análisis factorial exploratorio de la subescala de importancia en el cuestionario disciente parte de los 22 ítems que la integran. Suprimidos dos de ellos por causas psicométricas, la solución adoptada ordena los 20 ítems restantes en seis factores que explican el 61,36% de la varianza total. Como en el caso del profesorado, el método empleado para la extracción de dichos factores ha sido el análisis de componentes principales, siendo el criterio para la retención de los mismos la obtención de autovalores por encima del 1:

Tabla 75

Extracción de componentes principales y varianza total explicada en la subescala de importancia en estudiantes

	VARIANZA TOTAL EXPLICADA								
	Autovalores iniciales			Suma de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	TOTAL	% de la varianza	% acumul.	TOTAL	% de la varianza	% acumul.	TOTAL	% de la varianza	% acumul.
1	5,494	27,469	27,469	5,494	27,469	27,469	2,772	13,860	13,860
2	1,912	9,559	37,028	1,912	9,559	37,028	2,206	11,031	24,890
3	1,399	6,996	44,024	1,399	6,996	44,024	2,202	11,008	35,899
4	1,325	6,627	50,651	1,325	6,627	50,651	2,091	10,454	46,353
5	1,127	5,633	56,284	1,127	5,633	56,284	1,641	8,206	54,559
6	1,015	5,077	61,361	1,015	5,077	61,361	1,361	6,803	61,361

⁸⁴ Como en el caso del indicador C. 7.1., esta disposición inicial explica el lugar que ocupa en la solución final resultante, forzado a asociarse a indicadores dispuestos para otras dimensiones.

La solución presentada en la tabla 75 es la tercera de cuatro alternativas exploradas. Expuesta en el **ANEXO 21**, la solución inicial con los 22 ítems de la subescala de importancia daba cuenta de 6 factores que explicaban el 57,89% de la varianza total. Aunque coherente desde la perspectiva conceptual, el análisis de sus principales rasgos psicométricos, evidenciaba amenazas sobre la estructura interna del cuestionario. Su revisión daba cuenta de tres ítems *crossloaders*, es decir, elementos con pesos no pequeños (en torno a .30) en dos o más factores (Morales, 2013): el C.3.1., C.3.2. y E.4.1.3. Los factores deberían representar subgrupos de ítems mutuamente excluyentes, de modo que Prat y Doval (2005) recomiendan la supresión de aquellos ítems ambiguos, con saturaciones similares en más de uno. Dicha recomendación ha sido desarrollada por pasos, analizando los cambios producidos en la matriz de correlaciones:

- Partiendo de esta primera tentativa, se probó una *segunda solución* suprimiendo del análisis el ítem problemático de menor correlación, saber, el C.3.2. (*Diversificación de las condiciones de la evaluación*). El resultado fue un leve aumento de la proporción de varianza total explicada (59,73%) pero con persistencia de indicios de inadecuación psicométrica; el ítem C.3.1. seguía mostrando saturaciones similares en dos factores y para el indicador E.4.1.3. se producía un empeoramiento sustancial al aumentar simultáneamente su carga en tres de los seis factores. Sin embargo, suprimido este último ítem dentro de la misma solución, la proporción de varianza total explicada bajaba hasta el 55,45% y se evidenciaban nuevos casos de ítems ambiguos. En consecuencia, de este segundo análisis se ha adoptado la decisión de suprimir el ítem C.3.2. del cuestionario discente y de mantener provisionalmente el E.4.1.3.
- Partiendo de estas decisiones provisionales, se realizó un nuevo análisis suprimiendo el ítem C.3.1., tercera variable *crossloader* identificada en la solución inicial. Esta modificación supone extraer del cuestionario discente la dimensión tercera completa del modelo teórico (C.3.1. y C.3.2.), relativa a la atención a la diversidad en la evaluación. Genera una *tercera solución* en la que el porcentaje de varianza total explicada alcanza el 61,36%, el mejor resultado de todos los intentos. El análisis factorial apunta a la existencia de seis factores coherentes y con sentido desde la perspectiva teórica y conceptual. Los requisitos psicométricos necesarios para garantizar la claridad y solidez de la estructura interna del cuestionario siguen sin cumplirse en el caso del ítem E.4.1.3., que muestra saturaciones muy parecidas para tres factores diferentes. Sin embargo, suprimido dicho ítem de la solución, el resultado empeora sustancialmente: aparecen nuevos casos de elementos

ambiguos, se pierde sentido y claridad conceptual y la proporción de varianza total explicada desciende al 57%. Por todo ello, este tercer análisis concluye con dos decisiones provisionales:

- ✓ Suprimir del cuestionario discente los indicadores vinculados a la diversificación de la evaluación (dimensión tercera, ítems C.3.1. y C.3.2.) por causas psicométricas. Esta decisión está mediatizada por el propio procedimiento de análisis factorial, que identifica factores atendiendo a los indicadores con correlaciones altas o moderadas entre sí. Sin embargo, es preciso valorar el impacto que ha podido tener en este resultado la controversia natural que suscita la atención a la diversidad entre el alumnado (**ANEXO 22**). Como será desarrollado en próximos capítulos, la adecuación de los dispositivos evaluativos a las necesidades particulares de las/os estudiantes no representa una alternativa admitida y compartida de común acuerdo. A pesar de constituir un principio de calidad evaluativa señalado persistentemente desde la literatura específica (Ahumada, 2005; Cano et al., 2010; V. M. López et al., 2007), los datos recabados en esta muestra y en otras investigaciones similares sugieren que la evaluación correcta y justa sigue siendo interpretada en términos de homogeneidad y estereotipia, principalmente.
 - ✓ A pesar de la ambigüedad constatada, mantener de forma transitoria el ítem E.4.1.3. por dos razones: su supresión resiente la estructura interna desde el punto de vista psicométrico y desde la perspectiva conceptual el elemento está ubicado con sentido dentro de la escala (aparece agrupado junto a indicadores que representan la misma dimensión teórica de referencia).
- La cuarta alternativa de solución factorial analizada representa una comprobación final fundamentalmente. Aunque en todos los intentos la supresión tentativa del ítem E.4.1.3. empeoraba la claridad y solidez de la estructura interna, quedaba comprobar los cambios al suprimir del análisis sólo dicho elemento (manteniendo la dimensión tercera del modelo teórico con sus dos ítems). El resultado fue automáticamente desestimado ya que el porcentaje de varianza total explicada descendía hasta el 53,73%.

La tabla 76 recoge la matriz de componentes rotados correspondiente a la solución adoptada (la tercera). Suprimida la dimensión tercera, ordena los 20 ítems o variables restantes en seis factores identificados en la rotación ortogonal Varimax:

Tabla 76

Matriz de componentes rotados en la subescala de importancia en estudiantes

Ítem	Matriz de componentes rotados (Varimax con Kaiser)					
	COMPONENTE					
	1	2	3	4	5	6
C.7.1.	,434	-,181	,283	,332	-,032	,021
E.8.1.	,806	,166	,141	,033	,181	,108
E.8.2.	,849	,173	,045	,107	,196	,096
E.8.3.	,833	,245	,082	,024	,089	,086
E.4.2.1.	,217	,587	,011	,171	,027	,260
E.4.2.2.	,130	,789	,217	,042	,230	,035
E.4.2.3.	,128	,800	,162	,071	,202	-,022
C.1.1.	,028	,127	,613	,078	,061	,193
C.1.2.	,082	,150	,630	,160	-,060	,015
C.1.3.	,151	,049	,665	-,004	,142	-,089
C.1.4.	,007	,029	,683	,076	,118	,148
C.2.1.	,227	,125	,473	,364	-,063	-,044
C.5.3.	,026	,045	,070	,770	,161	,088
C.6.1.	,011	,033	,155	,786	,132	,030
C.6.2.	,140	,195	,099	,707	-,067	-,007
E.5.1.	,130	,156	,084	,101	,828	,123
E.5.2.	,221	,206	,113	,101	,804	,076
E.4.1.1.	,056	,015	,067	,013	,169	,860
E.4.1.2.	,321	,319	,173	,050	,059	,488
E.4.1.3.	,405	,425	,149	,163	-,091	,431

Nota: El color rojo en la fuente indica incumplimiento de alguno de los supuestos psicométricos barajados inicialmente para la valoración de la adecuación de la estructura interna.

Como en el caso del cuestionario docente, para valorar la adecuación y aceptabilidad de la estructura interna revelada, es preciso considerar indicaciones diferentes, sin tomar distancia de los propósitos de la investigación. Desde el punto de vista psicométrico caben tres apreciaciones principalmente:

- La carga de los ítems o su relevancia dentro de los factores delimitados es variable en la matriz de respuesta discente. Se detectan correlaciones moderadas (entre .40 y .59), altas (.60 y .79) o muy altas (por encima de .80), junto a cuatro componentes con saturaciones por debajo del medio punto⁸⁵. Aunque ello suponga el

⁸⁵ Dos de ellos proceden de las dimensiones segunda y séptima del modelo teórico, aquellas conformadas por un único indicador, forzado a asociarse a otros elementos al efectuar un análisis factorial.

incumplimiento de uno de los criterios psicométricos adoptados como referencia inicial, es preciso considerar el efecto del tamaño de la muestra en la interpretación de dichas cargas, ya que a mayor número de casos, mayor es también la probabilidad de obtener valores reducidos (Morales, 2013).

Frente a la muestra docente, integrada por 51 sujetos, la muestra de estudiantes está conformada por 704 casos, por lo que está justificado recurrir a otras referencias. A modo orientativo, Morales (2013) indica que para muestras de más de 400 sujetos, una carga de 0,30 puede considerarse aceptable, mientras que Costello y Osborne (2005) recomiendan no bajar de 0,35. Adoptando dichas estimaciones, puede afirmarse que la solución seleccionada responde a los requisitos de relevancia.

- El *número de componentes por factor* vuelve a representar una limitación psicométrica vinculada a la estructura externa del propio cuestionario. Un factor sólido debería estar conformado por al menos tres ítems, pero este requisito no se satisface en el quinto factor. El incumplimiento se contrarresta en las altas cargas de los dos elementos que concentra y su reseñable significación conceptual.
- En cuanto a la *claridad general de la estructura*, los ítems incluidos en el cuestionario del alumnado saturan un único factor sin dar muestras de ambigüedad. Este requisito se incumple de forma manifiesta en el ítem E.4.1.3., con saturaciones muy similares en tres de los seis factores. Aunque su supresión haya sido desestimada por empeorar la estructura psicométrica y conceptual de la solución, ofrece indicios de ambigüedad que deberán considerarse en la interpretación de los resultados a los que dé lugar.

Desde el punto de vista teórico, la estructura de factores presentada también resulta aceptable. Salvada la supresión de la dimensión tercera por causas psicométricas, relaciona con sentido los indicadores considerados y no distorsiona los propósitos de la investigación. Así se evidencia en la descripción de los seis factores resultantes, que muestra una concordancia razonable con las ocho dimensiones teóricas inicialmente previstas (ver figura 31):

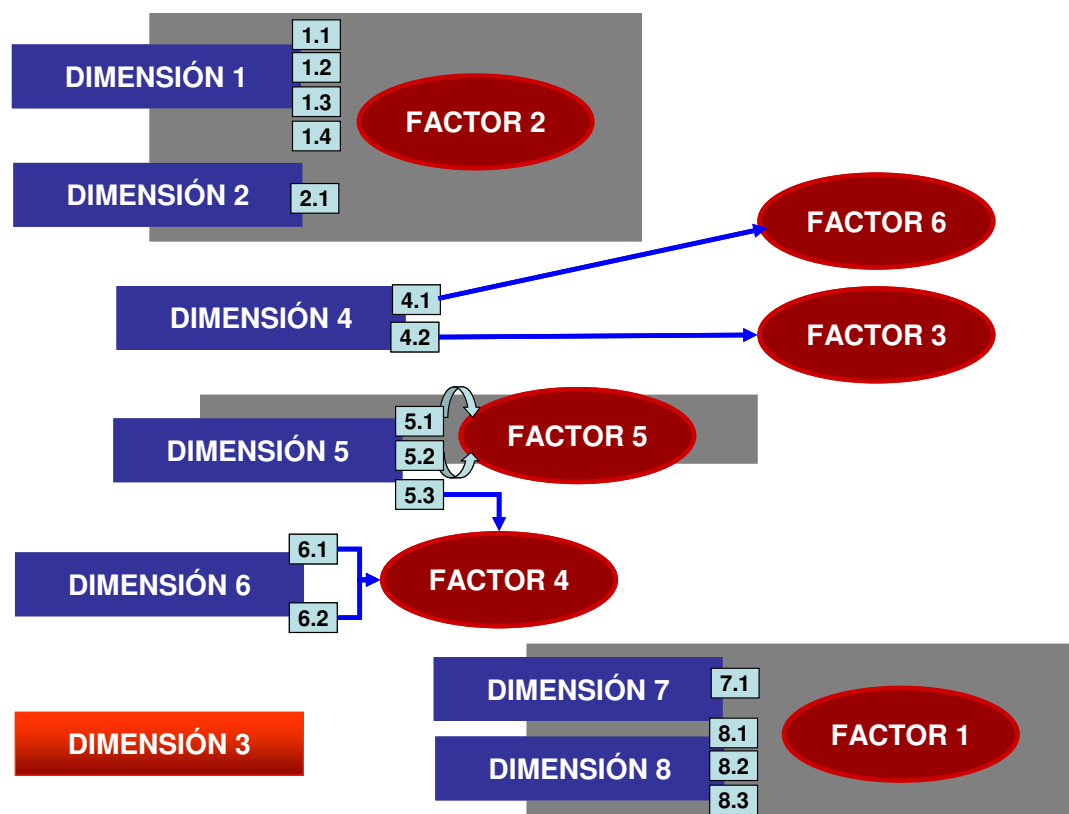


Figura 31. Estructura interna de la respuesta discente y relación con el modelo teórico de referencia.

Con respecto al modelo teórico de referencia el cambio más llamativo se debe a la desaparición de la dimensión tercera del cuestionario discente para garantizar el cumplimiento de los requisitos psicométricos asumidos. Las modificaciones restantes son menores y no alteran la significación conceptual:

- Cuatro de las dimensiones teóricas inicialmente previstas se reducen en dos factores. Las dimensiones relativas a la metaevaluación (octava) y a la planificación estratégica (primera) aparecen recogidas respectivamente dentro de los factores primero y segundo con todos sus componentes. Sin embargo, cada una de ellas se amplía por la fusión de las dimensiones teóricas conformadas por un único componente (la séptima y segunda). La clarificación de los criterios y condiciones de la evaluación (dimensión segunda) se vincula en la práctica a la planificación del sistema de evaluación con pleno sentido (factor segundo). La participación del alumnado en la evaluación (dimensión séptima) se agrupa con los tres componentes relativos a la metaevaluación, en los que también se valora la intervención de las/os estudiantes (en este caso, en la revisión del proceso).
- En otras dos dimensiones, se produce una leve reordenación por el cambio de ubicación de un elemento. El ítem C.5.3., integrado inicialmente en la dimensión

relativa al análisis e interpretación de evidencias (la quinta) pasa a formar parte del factor relacionado con la regulación del aprendizaje. Este cambio no altera el sentido o significado general de las dos dimensiones implicadas y replica la estructuración resultante en la muestra docente.

- Los dos factores restantes (el tercero y sexto) surgen de la descomposición de la dimensión cuarta, vinculada a la implementación del sistema evaluativo. Dicha dimensión se concreta en dos indicadores. El primero atiende al desarrollo coherente de los compromisos evaluativos y ha sido traducido en tres ítems en la escala de estudiantes (E.4.1.1., E.4.1.2. y E.4.1.3.). Bajo el denominador común del control de la puesta en práctica, dichos componentes pasan a constituir el factor sexto. El segundo indicador hace referencia a la gestión de los cambios producidos en la puesta en práctica del sistema evaluativo y también se presenta desglosado en tres ítems para su valoración por parte del alumnado (E.4.2.1., E.4.2.2. y E.4.2.3.). Dichos componentes conforman el factor tercero, constituido en torno a los ajustes de la evaluación.

La configuración de cada uno de los factores delimitados se concreta del siguiente modo:

Factor 1. Participación del alumnado en la evaluación.

Este primer factor hace referencia a la participación de las/os estudiantes en la evaluación del aprendizaje, ya sea en las fases de planificación, puesta en práctica (C.7.1.) o revisión (E.8.1., E.8.2. y E.8.3.). Surge de la fusión de dos de las dimensiones identificadas en la propuesta teórica presentada en el capítulo quinto: la relativa a la participación discente y la dedicada a la metaevaluación.

El factor lo componen cuatro ítems. Los tres primeros son específicos para estudiantes⁸⁶ y se corresponden con la octava dimensión del modelo teórico, relativa a la metaevaluación. En ellos se cuestiona la intervención discente en la revisión técnica, ética y pedagógica de la evaluación y obtienen correlaciones muy elevadas (ver tabla 77). Mucho menor resulta la saturación del ítem C.7.1., aunque desde la perspectiva conceptual su integración tenga pleno sentido. Relativo a la participación general del alumnado en la

⁸⁶ A ello responde que su codificación comience con una E (Estudiantes), según lo establecido en el ANEXO 17.

evaluación y común en ambas herramientas, representa el único componente de la dimensión séptima del modelo teórico. El peso de magnitud moderada que ocupa dentro del factor responde previsiblemente a esta circunstancia.

Tabla 77

Componentes del Factor 1 en la versión discente del cuestionario (Participación del alumnado en la evaluación) y sus cargas factoriales

Ítem	ENUNCIADO	Carga
E.8.2.	<i>Nos hace preguntas y pide nuestra opinión para valorar si su sistema de evaluación es riguroso, preciso y adecuado a la hora de valorar nuestro aprendizaje.</i>	.849
E.8.3.	<i>Nos hace preguntas y pide nuestra opinión para valorar si el sistema de evaluación que aplica es respetuoso con nuestros derechos y los principios éticos fundamentales (justicia, privacidad, bondad y transparencia).</i>	.833
E.8.1.	<i>Nos hace preguntas y pide nuestra opinión para valorar hasta qué punto el sistema de evaluación que aplica nos ayuda a aprender (si nos ayuda a centrar la atención, a mejorar, a trabajar a buen ritmo, etc.).</i>	.805
C.7.1.	<i>Desarrolla actividades que provocan la participación activa de las/os estudiantes en la evaluación.</i>	.435

Factor 2. Planificación del sistema evaluativo.

Como su propio nombre indica, el segundo factor hace referencia a la planificación del sistema evaluativo empleado en un módulo/asignatura determinado. Centra la atención en las características principales que determinan su configuración, así como en las acciones emprendidas para que el alumnado las conozca e interiorice a lo largo del proceso.

Este segundo factor está compuesto por cinco ítems comunes en ambas versiones del cuestionario y aglutina la primera y la segunda dimensión del modelo teórico de referencia. Los cuatro ítems con mayores cargas son los referidos a la planificación estratégica de la evaluación del aprendizaje (dimensión primera); con correlaciones altas, (entre .61 y .68), todos ellos aluden a decisiones ligadas al diseño evaluativo (ver tabla 78).

Como en el factor anterior, la diferencia de saturación contemplada en el ítem C.2.1. responde presumiblemente a su procedencia teórica (dimensión segunda, conformada por un único indicador). Plantea la interiorización de condiciones previstas en la planificación evaluativa, complementando con sentido los cuatro indicadores anteriores.

Tabla 78

Componentes del Factor 2 en la versión discente del cuestionario (Planificación del sistema evaluativo) y sus cargas factoriales

Ítem	ENUNCIADO	Carga
C.1.4.	<i>Descripción de las tareas de evaluación (demostración de contenidos o competencias).</i>	.682
C.1.3.	<i>Combinación de tareas evaluativas evaluación formativas y sumativas.</i>	.665
C.1.2.	<i>Coherencia/alineación de la evaluación con los objetivos formativos y la metodología de trabajo desarrollada.</i>	.630
C.1.1.	<i>Distribución de las tareas evaluativas a lo largo del proceso.</i>	.614
C.2.1.	<i>Desarrollo de actividades específicas para facilitar la interiorización y apropiación de los criterios de evaluación a lo largo del proceso formativo.</i>	.475

Factor 3. Ajustes del sistema evaluativo.

El tercer factor identificado se centra en la gestión de los ajustes que se efectúan en la práctica sobre los compromisos evaluativos adoptados. Se corresponde con el segundo indicador de la cuarta dimensión del modelo teórico, donde se aborda la implementación o puesta en práctica del sistema evaluativo.

Como refleja la tabla 79, el factor lo integran tres indicadores o ítems específicos para estudiantes. Los dos que hacen referencia a la comunicación y negociación de los cambios o ajustes producidos presentan correlaciones altas, ligeramente por debajo de .80. El ítem que plantea de forma general el cumplimiento de los compromisos evaluativos tiene un peso algo menor dentro del factor (moderado según las estimaciones barajadas).

Tabla 79

Componentes del Factor 3 en la versión discente del cuestionario (Ajustes en el sistema evaluativo) y sus cargas factoriales

Ítem	ENUNCIADO	Carga
E.4.2.3.	<i>Acuerda con nosotros los ajustes o modificaciones que introduce en el sistema evaluativo respecto a lo establecido en la documentación de la asignatura, en el caso de que se produzcan.</i>	.799
E.4.2.2.	<i>Comunica y justifica ante nosotros los ajustes o modificaciones que introduce en el sistema evaluativo respecto a lo establecido en la documentación de la asignatura, en el caso de que se produzcan.</i>	.785
E.4.2.1.	<i>Cumple los compromisos definidos sobre la evaluación de nuestro aprendizaje en la documentación de la asignatura (no modifica modalidades, plazos o condiciones sin justificación o aviso).</i>	.577

Factor 4. Regulación del aprendizaje por medio de la evaluación.

El cuarto factor hace referencia a la regulación del proceso de aprendizaje del alumnado a partir de los resultados puestos de manifiesto en la evaluación. Incluye elementos que hacen referencia a la comunicación de resultados y a los procesos de feedback y proacción o mejora que tienen lugar a partir de la misma.

El factor está compuesto por tres ítems con correlaciones altas, que también se encuentran en el cuestionario administrado al profesorado (ver tabla 80). Al igual que en la muestra docente, este factor une la dimensión sexta del modelo teórico (Regulación del aprendizaje) y el indicador de la dimensión quinta referido a la comunicación de resultados. Con pleno sentido, reafirma así uno de los ajustes que se realizará sobre el modelo de partida de cara a próximas indagaciones y profundización (**ANEXO 20**).

Tabla 80

Componentes del Factor 5 en la versión discente del cuestionario (Regulación del aprendizaje por medio de la evaluación) y sus cargas factoriales

Ítem	ENUNCIADO	Carga
C.6.1.	<i>Ofrece feedback a partir de las actividades evaluativas desarrolladas.</i>	.786
C.5.3.	<i>Presenta los resultados de forma clara, concisa y bien fundamentada, por medio de formatos que respetan la confidencialidad.</i>	.772
C.6.2.	<i>Especifica pautas claras de proacción para orientar a las/os estudiantes en el desarrollo de las competencias previstas.</i>	.703

Factor 5. Adecuación y precisión.

El quinto factor hace referencia a los rasgos psicométricos de la evaluación del aprendizaje, centrando la atención en el grado de la validez y fiabilidad que en la misma perciben las/os estudiantes.

El factor está conformado sólo por dos indicadores específicos del cuestionario discente, incumplimiento los requisitos generales de solidez. Sin embargo, esta limitación se contrarresta por las elevadas cargas que presentan dentro del mismo (ver tabla 81), así como por su significación conceptual. El primero de ellos plantea el tema de la adecuación o validez, entendida como la capacidad de la evaluación para centrarse y reflejar los aprendizajes esperados y no otras cuestiones relacionadas (E.5.1.). En paralelo, el segundo componente refiere a la precisión o fiabilidad en el establecimiento de distintos niveles de logro respecto a los mismos (E.5.2.).

Tabla 81

Componentes del Factor 5 en la versión discente del cuestionario (Adecuación y precisión) y sus cargas factoriales

Ítem	ENUNCIADO	Carga
E.5.1.	<i>El sistema de evaluación que aplica el/la profesor/a es capaz de determinar verdaderamente si un/a estudiante ha aprendido o no.</i>	.829
E.5.2.	<i>El sistema de evaluación que aplica el/la profesor/a es preciso y riguroso, capaz de diferenciar verdaderamente distintos niveles de logro o aprendizaje entre las/os estudiantes.</i>	.805

Factor 6. Control de las condiciones en las que se desarrolla la evaluación.

El sexto y último factor identificado hace referencia al control de las condiciones de la puesta en práctica del sistema de evaluación. Incluye tres ítems específicos del cuestionario del alumnado y cada uno de ellos hace referencia al control de un tipo de circunstancia con influencia potencial en la puesta en práctica del sistema evaluativo: el control de la legitimidad del proceso (E.4.1.1.), de la igualdad asociada a las condiciones prácticas (E.4.1.2.) y del ambiente en el que se desarrollan las tareas (E.4.1.3.).

El peso específico de los componentes dentro del factor es variable según contempla la tabla 82. La correlación es muy elevada en el ítem relacionado con las medidas para garantizar la legitimidad de la evaluación y tan sólo moderada para los que refieren a la igualdad de condiciones (E.4.1.2.) y a la seguridad del ambiente (E.4.1.3.). Como ya ha sido indicado, este último componente es el elemento más ambiguo de toda la escala, al correlacionar en similar medida con otros dos factores (el primero y en especial con el tercero, referente a los ajustes en el proceso evaluativo). A pesar de ello, su supresión ha sido descartada por empeorar el porcentaje de varianza total explicada y por la vinculación teórica que mantiene con los otros dos componentes del factor.

Tabla 82

Componentes del Factor 6 en la versión discente del cuestionario (Control de las condiciones en las que se desarrolla la evaluación) y sus cargas factoriales

Ítem	ENUNCIADO	Carga
E.4.1.1.	<i>Toma medidas para que las tareas evaluativas transcurran y se realicen sin riesgos que atenten a la legitimidad del proceso (plagios, suplantaciones, fraudes de cualquier tipo, etc.).</i>	.857
E.4.1.2.	<i>Toma medidas para favorecer que las tareas evaluativas transcurran en igualdad de condiciones, sin ventajas o desventajas asociadas al orden de realización, las fechas, el espacio o la modalidad de evaluación.</i>	.493
E.4.1.3.	<i>Toma medidas para que realicemos las tareas evaluativas en un ambiente favorable, de modo que nos sintamos cómodos y seguros durante su desarrollo (ambiente tranquilo, en condiciones físicas idóneas, etc.).</i>	.439

Ante los indicios de ambigüedad evidenciados, los resultados relativos a este último factor serán interpretados con cautela en los análisis practicados y sus conclusiones.

4.3 Fiabilidad.

Asociada a una amplia variedad de conceptualizaciones con matices diversos (estabilidad, reproductividad, consistencia, confiabilidad, etc.), la fiabilidad responde a la idea general del grado de precisión en la medida (Morales, 2007).

Desde el punto de vista práctico, puede considerarse la fiabilidad como la consistencia entre las puntuaciones que otorgan a una misma variable o evento diferentes evaluadores, o la persistencia de las puntuaciones cuando se aplica el mismo instrumento de evaluación en diferentes momentos. Por tanto, la fiabilidad puede entenderse como la exactitud de los datos en el sentido de su estabilidad, repetición, o precisión. (Albert, 2009, p. 102).

En la medida en la que pretendemos profundizar en el significado y las implicaciones de esta interpretación general, comienzan a hacerse manifiestas las distintas acepciones que adopta el término. Entre ellas, pueden distinguirse tres grandes tendencias (Albert, 2009):

- *La fiabilidad como predictibilidad o estabilidad en los resultados de la medición.* El fundamento de esta primera interpretación es la estabilidad de las respuestas de los sujetos, de modo que aplicaciones sucesivas de una misma herramienta deberían dar lugar a resultados coincidentes.
- *La fiabilidad como precisión o ausencia de distorsión en aquello que se mide.* Como señalan G. Pérez (1994) o Morales (2007) esta precisión puede ser independiente de la validez, es decir, cabe discriminar escrupulosamente distintos niveles en una variable que verdaderamente no se pretendía medir.
- *La fiabilidad como ausencia de error.* Esta tercera acepción define la fiabilidad en términos de precisión de medida, donde un mayor margen de error supondría una menor confiabilidad de los resultados obtenidos (Morales, 2007).

El significado de fiabilidad que mejor se ajusta a las necesidades y exigencias psicométricas de los propósitos de la presente investigación es la tercera, a saber, aquella que la define como ausencia de error y que se calcula por medio de coeficientes de *consistencia interna*. Esta consistencia refiere al grado de relación que existe entre los ítems que componen una escala (Prat & Doval, 2005). Frente a la homogeneidad conceptual en la que se apoya la validez de estructura, entendida como consistencia interna la fiabilidad ofrece una estimación de la homogeneidad empírica (Morales, 2007).

Este enfoque parte de la asunción de que en las diferentes puntuaciones de los sujetos de una muestra inciden de forma conjunta dos tipos de varianza: la verdadera, que se asocia a diferencias reales entre los casos y aquella otra asociada a la falta de relación de ítems que miden cosas distintas (Morales, 2007). Los coeficientes de consistencia interna expresan “hasta qué punto las respuestas son lo suficientemente coherentes (relacionadas entre sí) como para poder concluir que todos los ítems miden lo mismo y, por lo tanto son sumables en una puntuación total única que representa, mide un rasgo” (Morales, 2007, p. 6). Desde esta perspectiva, puede entenderse como consistente el ítem que correlaciona de forma positiva y moderada con el resto de elementos de la escala (Prat & Doval, 2005).

Por tanto, entendida como consistencia interna, la fiabilidad es el porcentaje de varianza verdadera contenida en la varianza total de los resultados de una muestra. A ella subyace una presunción de unidimensionalidad con respecto al cuestionario o la escala (Prat & Doval, 2005), porque de otro modo no cabría esperar que los diferentes ítems que lo conforman deban hacer referencia a un aspecto común. Como ya ha sido constatado al explorar la estructura interna de los cuestionarios docente y discente, ninguno de ellos se ajusta a este supuesto; cada uno de ellos incluye una subescala de importancia y otra de realización y ofrece mediciones en torno a seis factores, identificados en el análisis factorial exploratorio de las respuestas a la primera subescala. En consecuencia, la valoración de este requisito psicométrico se ha realizado de acuerdo a la estructura factorial asumida en cada caso.

Aunque existen distintas alternativas posibles, para valorar la consistencia interna de los resultados asociados a ambos cuestionarios se ha recurrido al coeficiente alfa de Cronbach. Éste devuelve una estimación de la consistencia interna en un intervalo de 0 a 1, donde el 0 indica una relación nula entre los ítems y el 1 una coincidencia perfecta. Los valores más deseables dentro de este intervalo son aquellos positivos y entre moderados y altos, supuestamente vinculados a ítems que miden un mismo atributo, aportando su propio matiz distintivo a la medición (Prat & Doval, 2005). A modo orientativo, Nunnally (1978)

establece que un alfa entre .75 y .90 sería lo aconsejable. Morales (2007) por su parte, matiza esta estimación y señala que la aceptabilidad de un determinado índice de fiabilidad depende de los propósitos a los que sirva la medición: cuando los resultados pretendan ser empleados para la toma de decisiones, la fiabilidad no debería bajar de 0,85, mientras que si la intención es contribuir a la investigación general en un tema o profundizar en la descripción de una situación dada y ofrecer feedback al respecto, pueden ser admisibles índices en torno a 0,60.

4.3.1 Consistencia interna del cuestionario del profesorado.

Para valorar la consistencia interna de los resultados registrados en la versión docente del cuestionario, se ha adoptado como referencia la estructura revelada al explorar su validez de contenido. Como resume la figura 32 se ha calculado el coeficiente alfa de Cronbach para cada uno de los seis factores y en las dos subescalas que contempla la herramienta: la de importancia y la de realización:

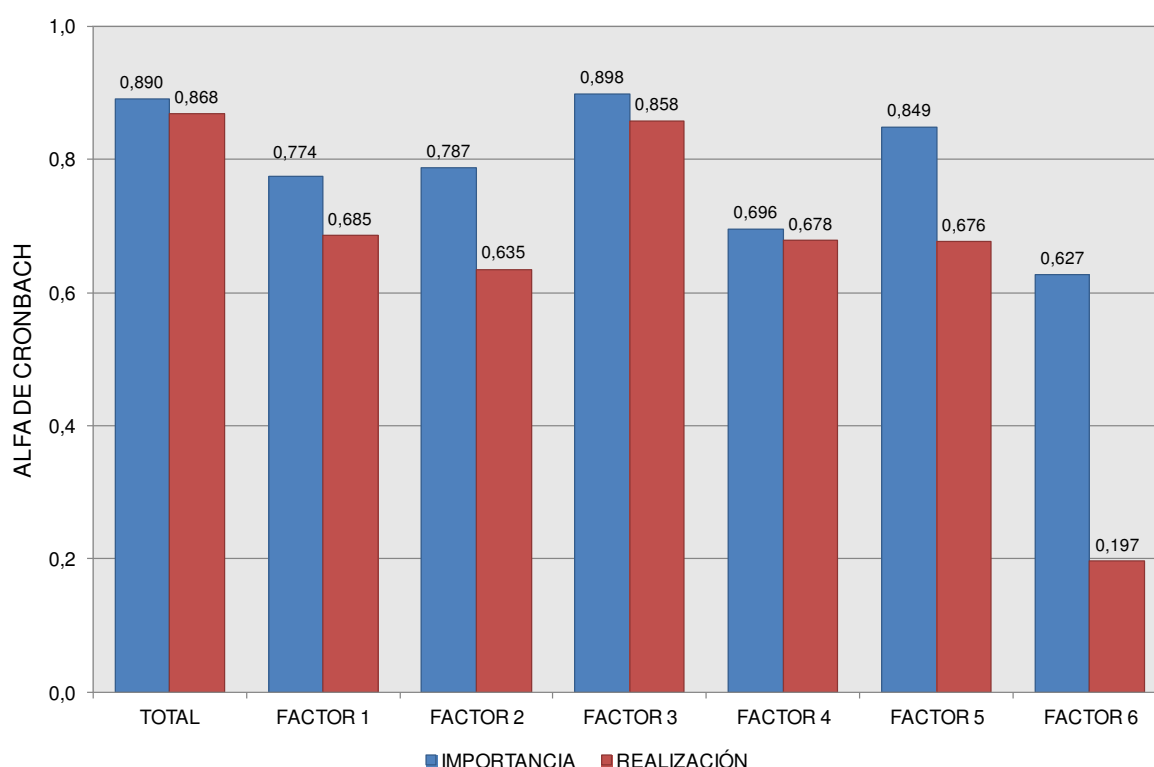


Figura 32. Índices de consistencia interna en los factores y subescalas del cuestionario docente. Los factores se presentan en el mismo orden que han sido definidos: 1. Planificación estratégica de la evaluación; 2. Regulación del aprendizaje y de la enseñanza por medio de la evaluación; 3. Metaevaluación; 4. Ajustes en la evaluación y participación discente; 5. Adecuación y precisión; 6. Puesta en práctica de la evaluación.

La consistencia interna total de ambas subescalas es muy elevada, con valores que superan claramente el .85. Esta tónica también se constata en la subescala de importancia, con resultados que avalan la fiabilidad en los seis factores discriminados. En todos ellos el alfa de Cronbach muestra niveles aceptables, con valores que oscilan entre el .89 registrado para el factor tercero (Metaevaluación) y el .62 del factor relativo a la Puesta en práctica de la evaluación (el sexto).

En la subescala de realización, la fiabilidad alcanza niveles satisfactorios en los cinco primeros factores. El tercer factor muestra una consistencia interna especialmente elevada, mientras que los cuatro factores restantes presentan valores razonables (entre .635 y .685) según las orientaciones de Prat y Doval (2005) o Morales (2007). No puede decirse lo mismo de la realización sujeta al sexto factor, para la que se estima una consistencia interna por debajo del .20 que no es aceptable desde el punto de vista psicométrico.

Siguiendo a Morales (2007), dos condiciones son necesarias para alcanzar un índice elevado de consistencia interna: la heterogeneidad de respuestas y la coherencia entre las mismas, condicionada a la relación de los componentes que conforman el factor. El bajo alfa de Cronbach obtenido al valorar el cumplimiento del factor sexto, parece relacionado con irregularidades en ambos sentidos, así como con la propia complejidad del objeto valorado:

- En cuanto a la relación de los componentes, el factor está conformado por dos únicos indicadores, inicialmente previstos para la valoración de dimensiones diferentes: el C.2.1., único elemento de la dimensión segunda referido a la clarificación de las condiciones de la evaluación, y el P.4.1., asociado a la implementación coherente de los sistemas de evaluación diseñados. La identificación de factores (y por extensión, la consistencia interna correspondiente a cada uno de ellos) fuerza el establecimiento de correlaciones entre los ítems o indicadores que conforman la escala, de modo que las dimensiones de único componente se ven abocadas a asociarse con otros elementos. En consecuencia, se producen asociaciones complejas, mientras que cuanto mayor sea el número de ítems por dimensión, mayor probabilidad existe de que la fiabilidad resultante sea alta (Morales, 2007).
- Unido a lo anterior se aprecia que, a pesar de que conceptualmente el factor tenga sentido, la correlación psicométrica entre sus dos componentes es muy baja (.011), igual que su covarianza o grado de variación conjunta (ver tabla 83). Este último

dato conecta con la heterogeneidad de las respuestas, segunda condición para la obtención de una alta consistencia interna. Sin diferencias entre las valoraciones que realizan los sujetos no se pueden hallar relaciones, ni variación conjunta (Morales, 2008a). Así sucede en la subescala de realización del factor sexto, donde el nivel de covarianza detectado sugiere respuestas muy similares en ambos ítems.

Tabla 83

Correlación y covariación de los componentes del factor sexto en la subescala de realización del cuestionario docente

SUBESCALA REALIZACIÓN FACTOR 6. <i>Puesta en práctica del sistema de evaluación</i>	CORRELACIÓN		COVARIANZA	
	C.2.1.	P.4.1.	C.2.1.	P.4.1.
C.2.1. Actividades para la interiorización y apropiación de criterios	1,000	,110	,804	,097
P.4.1. Implementación coherente del sistema evaluativo	,110	1,000	,097	,970

- Por último, la complejidad del propio objeto de valoración representa otro de los factores que puede incidir en el bajo índice de consistencia interna evidenciado (Morales, 2007). La competencia evaluativa difícilmente puede asumirse como un objeto unidimensional de análisis; incluso dentro de cada factor, parece razonable prever cierto grado de complejidad, que mitiga la asociación entre los ítems.

El balance general de los resultados presentados en la figura 32 lleva a dar por buena la consistencia interna de todas las subescalas implicadas en el cuestionario docente, a excepción de la referida a la realización del factor sexto. Considerando que los propósitos de la investigación son descriptivos y exploratorios y que la consistencia interna de la subescala de importancia de ese mismo factor es aceptable, se ha decidido mantenerlo de forma transitoria para no perder información. Sin embargo, los resultados asociados a dicha subescala sólo serán considerados a efectos descriptivos, sin que de los mismos se desprenda inferencia alguna.

4.3.2 Consistencia interna del cuestionario del alumnado.

Para determinar la fiabilidad de los resultados producidos a partir del cuestionario discente, se ha calculado la consistencia interna de cada uno de los factores detectados al analizar la validez de estructura, tanto en la estimación de importancia como en la de

realización. Según refleja la figura 33, la fiabilidad obtenida en los totales de ambas subescalas es muy aceptable, con coeficientes que superan el .85.

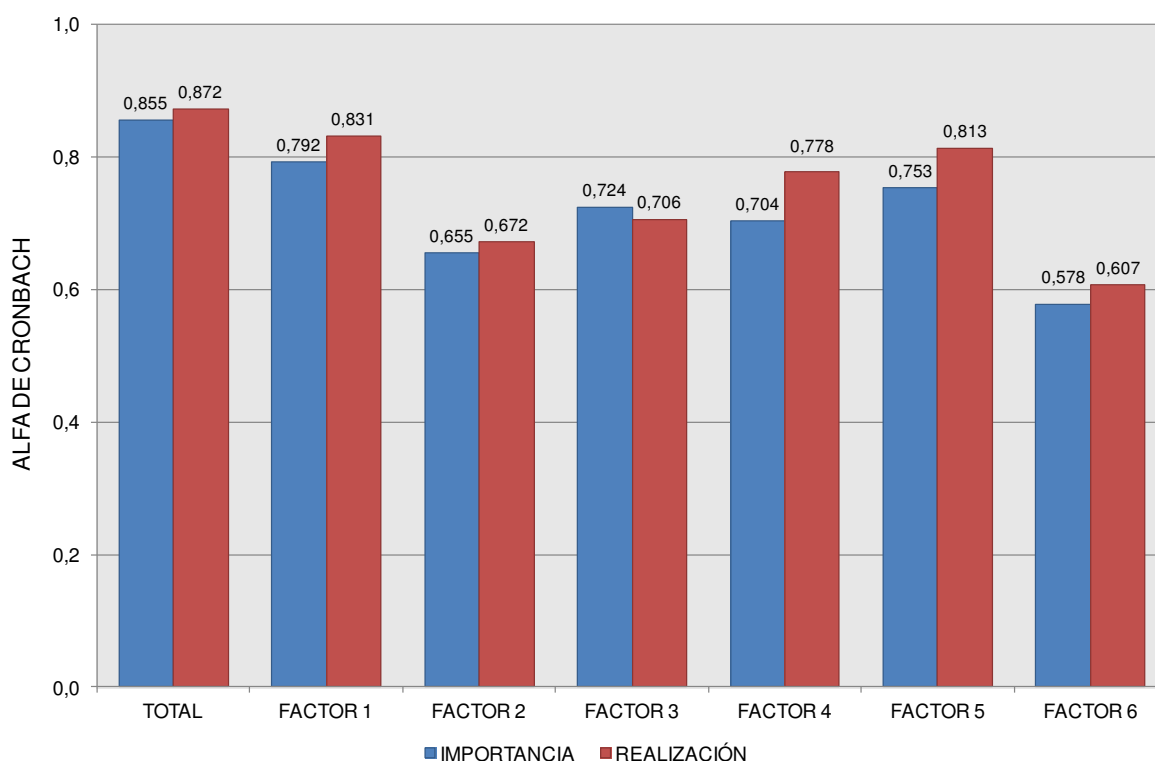


Figura 33. Índices de consistencia interna en los factores y subescalas del cuestionario discente. Los factores se presentan en el mismo orden que han sido definidos: 1. Participación del alumnado en la evaluación; 2. Planificación del sistema evaluativo; 3. Ajustes del sistema evaluativo; 4. Regulación del aprendizaje por medio de la evaluación; 5. Adecuación y precisión; 6. Control de las condiciones en las que se desarrolla la evaluación.

Los coeficientes representados para cada factor y subescala sugieren que la fiabilidad constatada es aceptable valorando los propósitos de la investigación en curso, exclusivamente formativos y orientados a la comprensión y la mejora. En términos generales, la subescala de realización o cumplimiento percibido, muestra una mayor consistencia interna que la de relevancia. Así se constata en cinco de los seis factores valorados, en todos salvo en el relativo a los ajustes en el proceso de evaluación (el tercero).

Con un alfa de .57 para la subescala de importancia y un .60 en la de realización, el factor sexto sobre el Control de las condiciones en las que se desarrolla la evaluación resulta el menos consistente de todos. En la interpretación de este resultado intervienen las mismas causas que ya han sido barajadas para explicar la escasa consistencia interna del sexto factor en la subescala de realización docente. Como en dicho caso, la correlación

media entre los tres componentes del factor es media-baja (.319) y se produce una cierta homogeneidad en las respuestas que generan (ver tabla 84). Aunque en este caso el coeficiente de consistencia obtenido sea mucho mejor, en su lectura es preciso sumar una cautela añadida, asociada a su estructuración interna. Como ya ha sido expuesto, uno de sus componentes (E. 4.1.3.) da muestras evidentes de ambigüedad al correlacionar simultáneamente con distintos factores (tabla 76).

Tabla 84
Correlación y covariación de los componentes del factor sexto en la subescala de importancia del cuestionario discente

SUBESCALA IMPORTANCIA			
FACTOR 6. <i>Control condiciones evaluación</i>		MEDIA	VARIANZA
Medias de los elementos		4,239	,059
Varianzas de los elementos		,598	,010
Covarianzas inter-elementos		,188	,000
Correlaciones inter-elementos		,319	,002

En el otro extremo, el factor primero, relativo a la Participación del alumnado en la evaluación, es el que mayor consistencia interna presenta, a la luz de los coeficientes reflejados en la figura 33 (ligeramente por debajo de .80 en la subescala de importancia y claramente por encima en la de realización). El factor quinto, sobre la Adecuación y precisión de la evaluación, es el segundo que mayor consistencia interna presenta, a pesar de estar formado únicamente por dos componentes.

4.4 Posibilidades y limitaciones asociadas a las herramientas de recogida de información diseñadas.

Este último apartado realiza un balance de las posibilidades y limitaciones que ofrecen los cuestionarios diseñados para efectuar la recogida de información en la muestra docente y discente. En ambos casos, las herramientas incluyen preguntas cerradas (en forma de ítem convencional o dispuestas en rúbrica) y preguntas abiertas. Las valoraciones contempladas en la tabla 85 derivan de las ventajas y desventajas asociadas a esta elección metodológica, así como de las propiedades psicométricas reportadas en los dos apartados anteriores.

Tabla 85

Valoración de las posibilidades y limitaciones asociadas a las herramientas de recogida de información

CRITERIO	POSIBILIDADES	LIMITACIONES
ALCANCE	• Eficiencia (buen equilibrio entre la cantidad de información recopilada y el coste proporcional).	• Superficialidad de la información recopilada. • La extensión del cuestionario puede ocasionar fatiga y desatención en ambas versiones. • Deseabilidad social, falta de objetividad (en especial, en el caso del profesorado).
PROFUNDIDAD	• Integración de preguntas abiertas para contrarrestar la superficialidad de las cuestiones cerradas.	• Riesgo de dispersión en las preguntas abiertas e incremento de los costes de registro y análisis.
UNIVOCIDAD	• En la versión discente, alta tasa de respuesta y uniformidad de interpretación gracias a la aplicación guiada.	• En la versión docente, riesgo de interpretaciones desiguales para los indicadores y descriptores por la aplicación autónoma.
INTEGRACIÓN DE RÚBRICAS	• Claridad que aportan las rúbricas de evaluación como recurso para unificar interpretaciones de los distintos grados de realización. • Valor formativo de las rúbricas al dibujar los logros deseables en la evaluación del aprendizaje.	• Dificultades para captar la amplia casuística que se constata en la realidad en una rúbrica (rudeza de medida y rigidez del instrumento frente a contextos singulares). • Posibles dificultades en el empleo autónomo de la rúbrica por parte del profesorado.
VALIDEZ Y FIABILIDAD	• Niveles razonables de consistencia interna en los seis factores de ambos cuestionarios (salvo en la realización del factor sexto docente). • En general, una estructuración interna clara de las respuestas a ambos cuestionarios (salvo en el sexto factor de la versión discente).	• Baja fiabilidad en la subescala de realización del factor sexto del cuestionario docente (empleado sólo con fines descriptivos). • Indicios de ambigüedad en la estructuración interna del factor sexto del cuestionario discente (interpretado con reservas y cautela). • Las dimensiones teóricas basadas en un único componente producen un efecto negativo en la claridad y solidez de la estructura interna de las herramientas y, en consecuencia, de la consistencia interna.

Entre las principales ventajas y potencialidades de los instrumentos utilizados, destaca en primer lugar su eficiencia. Los cuestionarios diseñados para estudiantes y profesorado permiten recopilar una cantidad de información muy importante con un coste relativamente razonable, sobre todo, si se compara con el que generan otras técnicas como la entrevista. En contrapartida, se asumen como inconvenientes parejos la superficialidad de las respuestas, la potencial falta de sinceridad o las dificultades asociadas a la adecuación léxica (Albert, 2009; Torrado, 2009). Estos riesgos se hacen especialmente manifiestos en el cuestionario docente, ya que la aplicación representa una autoevaluación,

donde el riesgo por falta de objetividad se hace mayor aún (B. García et al., 2012; Saroyan, 2001). Por otra parte, su aplicación también es autónoma, de modo que las respuestas se plantean desde la particular interpretación que realice el sujeto. En el caso del alumnado, este riesgo de interpretación dispar de significados se ha contrarrestado por medio de un procedimiento de recogida dirigido, donde se ha detallado el sentido de cada indicador y los descriptores de realización que lo acompañan. Aunque haya supuesto un incremento del coste de la investigación, este procedimiento ha favorecido a su vez, una tasa de respuesta muy elevada por parte del alumnado, difícil de garantizar en procesos de cumplimentación autónoma (on-line o en papel).

La cantidad de información recopilada ha sido reseñada como ventaja asociada a ambas herramientas, sin embargo ello redundante en la extensión de las mismas y en los riesgos consiguientes de desatención y fatiga por parte de quien responde. Este riesgo fue indicado en la aplicación piloto de ambas versiones y ha sido objeto de ardua deliberación en el proceso de diseño de las herramientas definitivas; se ha tratado de mantener el equilibrio entre simplificar en exceso las herramientas minimizando el número de indicadores consultados y superar el límite de lo abarcable en la práctica. Como ya ha sido detallado, este intento de reducción ha supuesto la adopción de ciertos riesgos psicométricos (dimensiones con uno y dos indicadores en varios casos), pero aun así, la herramienta exige una cumplimentación más trabajosa que cualquiera de las que han sido consultadas como referencia.

Para hacer frente a la superficialidad de la información devuelta en los cuestionarios, se han integrado unas preguntas finales en modalidad abierta y/o semiabierta. Este tipo de cuestiones aportan como ventaja principal la obtención de respuestas directas por parte de los agentes encuestados, sin ceñirse a una serie de alternativas predeterminadas de antemano e incidiendo en elementos realmente significativos para ellos. Como señala Albert (2009) esta amplitud en ocasiones lleva a respuestas dispersas que exceden el objeto de estudio. Además, este tipo de cuestiones conlleva un trabajo de registro y análisis más complejo y costoso que el de las preguntas cerradas.

Uno de los rasgos que más diferencia los cuestionarios elaborados de otras herramientas consultadas es el empleo de rúbricas o matrices de evaluación. Las rúbricas centran la atención de quien las cumplimenta, a la vez que definen los logros deseables para cada caso (Andrade, 2000, 2005; Blanco, 2008b; Reddy & Andrade, 2010). Incrementan la transparencia y reducen la ansiedad en la valoración, subrayando su potencialidad formativa (Atkinson & Lim, 2013; Panadero & Jonsson, 2013). Desde esta perspectiva, además de contribuir al establecimiento de un criterio común para efectuar las valoraciones, la propia cumplimentación de la herramienta representa para docentes y

discentes una experiencia formativa, en la medida en la que detalla los rasgos que reúnen las prácticas evaluativas deseables. Este beneficio ha sido especialmente constatado por parte del profesorado participante que, en el estudio piloto y en comentarios informales, ha indicado el valor de la plantilla de evaluación como guía para la revisión y redefinición de los sistemas evaluativos. Sin embargo, el empleo de rúbricas también implica la asunción de ciertos riesgos o limitaciones. La primera de ellas está ligada a la propia configuración de la plantilla de valoración y la segunda al uso que de la misma se realice:

- La claridad de la rúbrica ha sido señalada como ventaja, al aportar un criterio para unificar las diferentes interpretaciones que puedan generarse en torno al grado de realización de los indicadores. Sin embargo, considerando la complejidad y singularidad propia de los contextos educativos, también puede resultar un instrumento rígido que no termine de captar lo más relevante de cada situación. Detallar cinco niveles de realización progresivos para cada indicador evaluado, ha representado un trabajo concienzudo y muy laborioso. En cada indicador se ha tratado de captar toda la casuística previsible, pero es difícil responder con claridad a todas las situaciones. A la hora de responder, estudiantes y profesorado también han manifestado cierta dificultad para posicionarse en un descriptor u otro, indicando que en ocasiones la realidad se sitúa en puntos intermedios. Esta limitación refleja lo que Siegel (1975) califica como la rudeza del instrumento de medida, que representa pobremente la continuidad real que existe en las variables dependientes medidas (relevancia y realización).

Para poder responder a esta limitación, sería interesante poder llevar a cabo entrevistas, de modo que la rúbrica represente el marco común de referencia y sobre la misma se puedan efectuar las matizaciones y comentarios oportunos. En contrapartida, ello supondría asumir un incremento del coste de la investigación y una amenaza al derecho de anonimato de las/os participantes.

- La segunda limitación está relacionada con el propio uso de las rúbricas. Como señalan Hafner y Hafner (2003), Andrade (2000, 2005) o Reddy y Andrade (2010), el alumnado necesita apoyo para trabajar con las matrices de evaluación. En su opinión no debe darse por supuesto que una rejilla de valoración represente un recurso suficientemente sólido para orientar por sí mismo la evaluación, a pesar de la claridad aparente. Así lo han constatado en un estudio reciente Panadero y Jonsson (2013), al detectar que la efectividad del uso de las rúbricas en manos discentes aparece asociada a varios factores (nivel educativo, sexo u objeto de evaluación, entre otros posibles). En consecuencia, para la cumplimentación de los

cuestionarios en la muestra de estudiantes se han organizado sesiones dirigidas, destinadas a unificar interpretaciones y brindar el apoyo necesario. En el caso del profesorado, la respuesta autónoma del cuestionario no evita estos riesgos, de modo que incrementa el riesgo de imprecisión, disparidad de interpretación y sesgo por deseabilidad social.

Para finalizar esta síntesis de posibilidades y limitaciones vale la pena reparar mínimamente sobre los rasgos psicométricos que han mostrado los cuestionarios en su aplicación. Dentro de las limitaciones referidas en los anteriores apartados, las herramientas empleadas han puesto de manifiesto buenos niveles de validez de estructura y de consistencia interna en su administración en la muestra docente y discente. Las limitaciones observadas se centran en dos valoraciones precisas, en las que se debe adoptar una especial precaución en el uso e interpretación. Presumiblemente, ambas representan la consecuencia de la acotación deliberada del número de indicadores en el modelo teórico de referencia:

- La realización vinculada al factor sexto del cuestionario docente, que sólo será empleado con fines descriptivos por la baja fiabilidad que manifiesta su índice de consistencia interna.
- Las respuestas asociadas al sexto factor del cuestionario dirigido al alumnado, que en su estructuración interna incluye un elemento ambiguo que correlaciona con dos factores más.

Las consideraciones expuestas en este apartado sobre los cuestionarios diseñados, así como los resultados y valoraciones derivados de su primera aplicación en la presente investigación, se han concretado en una revisión de la rúbrica original que desglosa la propuesta teórica sobre la competencia evaluativa. El resultado se presenta en el **ANEXO 20**, que debería ser considerado en la adecuación de nuevas herramientas.

5. Recogida de información

Los datos en los que se apoyan las conclusiones empíricas de la investigación en curso han sido recopilados en dos fases como detalla la tabla 86:

Tabla 86
Síntesis del proceso de recogida de información desarrollado

DESCRIPCIÓN	FASE I	FASE II
FECHA	Segundo cuatrimestre del curso 2011/2012.	Primer cuatrimestre del curso 2012/2013.
MOMENTO	Dos últimas semanas de abril y primera semana de mayo, aproximadamente a dos semanas del final del cuatrimestre.	Última semana de noviembre y primera quincena de diciembre, aproximadamente a dos semanas del final del cuatrimestre.
DESTINATARIOS POTENCIALES	Población acotada en las primeras fases del muestreo: - Profesorado vinculado a los Grados de Educación Infantil y/o Primaria en los dos centros (UD y BAM) y con docencia adjudicada en el momento de la consulta. - Alumnado matriculado en los Grados de Educación Infantil y/o Primaria en ambos centros en el momento de la consulta.	Excluidos las/os participantes en la fase anterior de recogida: - Profesorado vinculado a los Grados de Educación Infantil y/o Primaria en ambos centros (UD y BAM) con docencia adjudicada en el momento de la consulta. - Alumnado matriculado en 1º, 2º ó 3º de los Grados de Educación Infantil y/o Primaria en ambos centros en el momento de la consulta.
MUESTRA DOCENTE CONSULTADA	38 docentes.	17 docentes.
MUESTRA DISCENTE CONSULTA	342 estudiantes (13 sesiones de recogida en el mismo número de grupos o conglomerados).	362 estudiantes (7 sesiones de recogida en el mismo número de conglomerados).
PROCEDIMIENTO	- DOCENTES. Respuesta autónoma y no presencial del cuestionario de autoevaluación. Salvo en tres casos, la cumplimentación se ha efectuado en papel. Devolución por medio de un procedimiento que salvaguarda el anonimato (codificación de la identidad y entrega en sobre cerrado en un lugar pactado). - ESTUDIANTES. Respuesta anónima, presencial y dirigida en horario lectivo. Distribución aleatoria de cuestionarios con el nombre del módulo/ asignatura en el que basar la valoración, según confirmación previa de participación docente. Devolución anónima in situ una vez finalizada la sesión.	

Como ya ha sido anticipado, la recogida de información en el colectivo docente se ha llevado a cabo por medio de un procedimiento de aplicación autónoma, que ha simplificado sustancialmente su desarrollo. No así en la muestra discente, donde valorada la complejidad de la herramienta diseñada, la recopilación se ha realizado en una sesión de evaluación presencial y dirigida.

Previo acuerdo con el profesorado participante en el estudio, por cada grupo de estudiantes o conglomerado incluido en la muestra se ha previsto un espacio de trabajo de 45 minutos para aplicar los cuestionarios⁸⁷. Para evitar diferencias, todas las sesiones han sido dirigidas por la doctoranda siguiendo el mismo protocolo:

⁸⁷ Como se ha declarado en el apartado de las variables independientes, en dos casos la aplicación de este protocolo se tuvo que reducir en 15 minutos aproximadamente. En los grupos afectados, los datos de

- *Presentación-sensibilización.* Introducción inicial al propósito de la investigación y confirmación de la colaboración discente.
- *Introducción.* Presentación de los módulos/asignaturas a valorar en un mismo grupo y referencia a las/os profesoras/es responsables en cada caso⁸⁸. Exposición de las pautas para garantizar el anonimato de las respuestas (distribución aleatoria de cuestionarios con el nombre del módulo/asignatura detallado e importancia de no revelarlo durante el proceso de aplicación).
- *Cumplimentación guiada de los datos de identificación.* Presentación de las preguntas y respuesta autónoma y anónima por parte del alumnado.
- *Cumplimentación guiada de las rúbricas de evaluación e ítems.* En cada uno de los indicadores evaluados, exposición del significado general de lo que pretende valorarse y de los cinco niveles de realización detallados. Dicha exposición la ha realizado la persona encargada de la recogida, mientras que cada estudiante iba tomando nota anónimamente de sus respuestas en el cuadernillo.
- *Entrega de los cuestionarios cumplimentados.* Devolución de los cuestionarios cumplimentados. La entrega se ha realizado de forma conjunta (por filas) y con el cuadernillo cerrado para proteger el anonimato de las/os participantes.

En total, para acceder a los 704 sujetos que conforman la muestra discente se han realizado sesiones de cumplimentación dirigida en 20 grupos o conglomerados de estudiantes de primer, segundo o tercer año de ambos centros y titulaciones. En ambas fases de la recogida la principal dificultad ha venido asociada a la accesibilidad. Para garantizar que el alumnado disponga de datos suficientes para valorar el sistema evaluativo aplicado por un determinado docente en el módulo/asignatura de referencia, la administración guiada de los cuestionarios no podía llevarse a cabo en la primera mitad del cuatrimestre, donde el acceso al horario lectivo hubiera resultado más sencillo.

identificación y las cuestiones vinculadas a las dimensiones 1, 2, 3 y 5 han sido cumplimentadas de forma dirigida, y de manera autónoma el resto de las preguntas. Esta modalidad de aplicación también se ha desarrollado íntegramente de forma presencial y ha finalizado con la entrega de los cuestionarios.

⁸⁸ La información proporcionada depende directamente de la afijación de conglomerados que haya sido definida en el procedimiento de muestreo (**ANEXO 12**).

6. Previsión y desarrollo de los procedimientos de análisis

Los apartados recorridos hasta el momento han cuenta del plan diseñado y puesto en práctica para recopilar de estudiantes y profesorado información que permita responder a los cuatro objetivos que guían el marco empírico (tabla 57). Este nuevo apartado, describe los análisis que se han practicado sobre la misma para llegar a los resultados y las conclusiones presentadas en el próximo capítulo.

La investigación en curso integra en su desarrollo dos grandes tipos de análisis: el análisis cuantitativo, centrado en relaciones y magnitudes de las variables numéricas consideradas, y el análisis cualitativo, centrado en el registro y categorización de las respuestas ofrecidas por profesorado y alumnado a las cuestiones abiertas. La presencia e incidencia de ambos no es equitativa; el grueso del análisis efectuado es cuantitativo, mientras que la parte cualitativa adopta un carácter más bien complementario.

6.1 Análisis cuantitativos previstos para la consecución de los objetivos.

Los análisis cuantitativos se han realizado sobre la información recopilada en las preguntas cerradas de los cuestionarios docente y discente y las variables de identificación. Dependiendo de su cometido o finalidad principal, dichos análisis se han desarrollado por medio de procedimientos estadísticos descriptivos o inferenciales. Para su realización, se ha recurrido a uno de los paquetes estadísticos más extendidos y reconocidos en la investigación social, a saber, el SPSS (Statistical Package for Social Science).

Como su propio nombre indica, los **estadísticos descriptivos** responden al cometido general de describir las respuestas ofrecidas por las/os docentes y estudiantes que integran la muestra. Incluyen toda una serie de análisis univariable que contribuyen a resumir las valoraciones realizadas por cada colectivo, así como a cuantificar sus propios rasgos definitorios (variables socio-demográficas, académico-profesionales, o institucionales). Las más presentes son la distribución de frecuencias, las medidas de tendencia central (promedios, moda, mediana) o las de dispersión (desviación típica y varianza). Concretamente, este tipo de análisis ha sido empleado para describir la muestra y para responder a los objetivos específicos E1 y F1 (tabla 57).

Junto a los objetivos descriptivos reseñados, la presente investigación incluye otro tipo de propósitos que apuntan al establecimiento de relaciones o diferencias generalizables a la población de referencia. Son los objetivos específicos E2, F2 y G, cuya resolución exige la realización de **inferencias estadísticas bivariable**.

Inferir es sacar consecuencias (Real Academia Española, 2014) y la estadística proporciona herramientas útiles para ello. La investigación realizada en ciencias sociales emplea fundamentalmente dos tipos de técnicas de inferencia estadística: las paramétricas y las no paramétricas. Ambas parten de rasgos y supuestos de aplicación diferentes, de modo que la elección debe partir de su consideración cautelosa.

Las **técnicas estadísticas paramétricas** son aquellas que se basan en suposiciones acerca de los parámetros de la población de la que se ha extraído la muestra, de ahí su denominación. Su aplicación requiere el cumplimiento de una serie de requisitos previos que incluyen los supuestos de aplicación derivados del modelo estadístico de la prueba y condiciones relativas a la escala de medida aplicada para las variables dependientes. Resumidas en la tabla 87, las pruebas de inferencia estadística paramétricas parten de tales asunciones, independientemente de que se realice su comprobación o no.

Tabla 87
Supuestos de aplicación de las técnicas estadísticas paramétricas

CONDICIÓN	DESCRIPCIÓN
Independencia de las observaciones o medidas	Las pruebas paramétricas deben efectuarse sobre observaciones independientes entre sí. La selección de casos dentro de la población debe ser independiente y también la de puntajes.
Normalidad	La distribución de la variable dependiente debe ajustarse a la distribución normal, al menos, en la población de referencia.
Homogeneidad de las varianzas	La varianza de la variable dependiente debe ser homogénea en los grupos que establezca la variable independiente (homocedasticidad). El objetivo es conseguir una dispersión por igual de la varianza de la variable dependiente a lo largo del rango de los valores de la variable independiente, evitando la perturbación conocida como heterocedasticidad.
Linealidad	Este requisito es aplicable sólo cuando se pretende realizar un análisis de varianza para la comparación de medias entre tres o más grupos. Indica que el modelo a contrastar predice los valores de la variable dependiente analizada siempre que se produzca una modificación en las variables independientes.
Fortaleza de medida	Las variables dependientes deben estar medidas al menos en una escala de intervalo (y preferiblemente continua), de modo que sea posible realizar operaciones aritméticas con los puntajes. Las escalas ordinales o nominales no se ajustan al sistema numérico aritmético en el que se basan las pruebas paramétricas.

CONDICIÓN	DESCRIPCIÓN
Tamaño de la muestra	Existe cierto acuerdo en que el tamaño de la muestra condiciona el empleo de pruebas estadísticas paramétricas. En muestras pequeñas es más probable que se incumpla el criterio de normalidad. En consecuencia, 30 suele señalarse como el número mínimo de casos a incluir en la muestra.
Tipo de muestreo	La inferencia estadística que ofrecen las pruebas paramétricas se basa en probabilidades. Su uso es desaconsejable en muestras conformadas por medio de procedimientos no aleatorios.

Nota: Síntesis elaborada a partir de R. Álvarez (1995), Berlanga y Rubio (2012), Iglesias y Sulé (2005), Manzano y Braña (2005), Rubio y Berlanga (2012), Siegel (1975) y Vilá y Bisquerra (2009).

Los supuestos de aplicación expuestos son amplios y exigentes, por ese motivo, si se satisfacen, las pruebas paramétricas son muy potentes. Frente a ellas, las pruebas de distribución libre o **técnicas estadísticas no paramétricas** “engloban una serie de pruebas estadísticas que tienen como denominador común la ausencia de asunciones acerca de la ley de probabilidad que sigue la población de la que ha sido extraída la muestra” (Berlanga & Rubio, 2012, p. 101). Una prueba no paramétrica es aquella cuyo modelo no especifica condiciones sobre los parámetros de la población de la que sale la muestra (Siegel, 1975).

En coherencia con la definición ofrecida, los supuestos de aplicación de las pruebas estadísticas no paramétricas tienden a definirse en negativo, es decir, señalando aquello que no se requiere para su uso. Son aplicables al análisis de variables dependientes que no se ajustan a la distribución normal, cuando existe heterocedasticidad constatada entre las varianzas de los grupos comparados, en muestras pequeñas e incluso con escalas de medición nominales y/u ordinales.

6.1.1 Entre la robustez y la cautela: el debate sobre la elección de procedimientos paramétricos y no paramétricos y la postura adoptada.

De acuerdo con Siegel (1975), la elección de una familia de técnicas u otra, debe depender de la concienzuda consideración de los siguientes aspectos:

- La **potencia de la prueba**, determinada por su capacidad potencial para captar diferencias o correlaciones significativas verdaderas y detectar aquellas que son falsas.
- El **modelo estadístico o las suposiciones de la prueba**, que incluye consideraciones sobre el modo en que se obtenga la muestra de puntajes y la

naturaleza de la población de la que procede la muestra. Cada técnica de inferencia estadística está sujeta a una serie de supuestos, que serán más amplios y exigentes, cuanto mayor sea la potencia de la prueba. El incumplimiento de los supuestos de aplicación de una prueba pone en entredicho su potencia, de modo que la elección de una técnica u otra, debe también reparar sobre los mismos.

- La **escala o el tipo de medida** que se haya empleado para operativizar las variables cuya estimación se pretende.

Sin embargo, la revisión de la literatura deja entrever un encendido debate sobre el rigor que cabe aplicar en la lectura de tales condiciones y las consecuencias de no realizar una elección estricta. En síntesis, podemos diferenciar dos grandes posturas en torno a esta cuestión: una primera que apela a la robustez de las pruebas paramétricas y una segunda más conservadora.

De un lado aparecen testimonios que sugieren la *conveniencia de priorizar el empleo de pruebas paramétricas, incluso cuando no se satisfagan todos los supuestos de aplicación descritos* para las mismas. Quienes defienden esta postura, confían en la potencia de las mismas y en su robustez, calificativo empleado para indicar la capacidad de violar los supuestos de una prueba estadística sin repercutir sustancialmente en las conclusiones.

Se ha observado que en la práctica muchos supuestos paramétricos no se cumplen *strictu sensu*, pero casi se cumplen. Por eso, aunque no se confirmen estrictamente todos los supuestos paramétricos, se pueden aplicar pruebas paramétricas, conociendo los efectos que ello pueda suponer sobre los resultados. Desde diversas perspectivas se aconseja el uso de pruebas paramétricas siempre que sea posible, ya que hay razones teóricas y empíricas que justifican que muchas pruebas paramétricas son lo suficientemente robustas para que los resultados no se vean seriamente alterados por ligeras desviaciones de los supuestos paramétricos. (Vila & Bisquerra, 2009, pp. 267-268).

Desde esta primera perspectiva, el empleo de pruebas no paramétricas se recomienda sólo en aquellos casos en los que los supuestos paramétricos se incumplan ostensiblemente. Sin embargo, existe una segunda tendencia de opinión que afirma que *el empleo de pruebas paramétricas en condiciones indebidas puede llevar a cometer*

importantes errores estadísticos (R. Álvarez, 1995; Siegel, 1975). Se trata de una postura más conservadora, que sostiene la importancia de plantear mediciones consistentes con el modelo estadístico asociado al tipo de prueba aplicada. Cuando estos supuestos no se cumplen, es muy difícil valorar la potencia de una prueba, incurriendo en riesgos incontrolados.

Desde esta perspectiva conservadora se aportan varias objeciones al empleo generalizado e inconsciente de los procedimientos paramétricos:

- Comprometiendo la potencia de las pruebas aplicadas, a menudo no se realiza una comprobación explícita del requisito de normalidad o simplemente no es posible efectuarla (Siegel, 1975).
- Aplicar estadística paramétrica con una escala de medición más débil que el intervalo puede introducir distorsiones importantes en la interpretación de los resultados.

El científico social, que toma la física como modelo, suele intentar algo parecido en la calificación o medición de las variables sociales; pero, en sus escalas, muy a menudo menosprecia un fundamento de la teoría de la medición. Pasa por alto que, para hacer ciertas operaciones con los números que ha asignado a las observaciones, la estructura del método de correspondencia de los números (puntajes) a las observaciones debe ser isomórfica con respecto a alguna estructura numérica que incluya estas operaciones. (Siegel, 1975, p. 42).

- El paquete estadístico SPSS incorpora procedimientos para realizar los análisis asumiendo la existencia de varianzas homogéneas o heterogéneas entre los grupos comparados. Sin embargo, Milligan, Wong y Thompson (1987) concluyeron que el modelo ANOVA proporciona resultados severamente sesgados cuando los datos incumplen el supuesto previo de homogeneidad de las varianzas.

Consideradas ambas posturas y sus respectivas aportaciones, en la investigación en curso se ha decidido adoptar como referencia *elemental la estadística no paramétrica* para la realización de las inferencias previstas en los objetivos. Dicha decisión ha sido adoptada en consideración de las siguientes circunstancias:

- Realizado un primer análisis exploratorio, se constata que el *criterio de normalidad* no se cumple en la muestra y tampoco disponemos de datos para suponer que esté presente en la población de referencia. La tabla 88 recoge los resultados de la simulación efectuada, según los análisis previstos en los tres objetivos que implican inferencia estadística:

Tabla 88

Análisis tentativo para la valoración del supuesto de normalidad en la muestra

	Nº ANÁLISIS	DISTRIBUCIÓN NORMAL	FUERA DE NORMALIDAD
OBJETIVO E2	294	24,14%	75,86%
OBJETIVO F2	196	2,55%	97,44%
OBJETIVO G	14	0%	100%

- En vista de los resultados puestos de manifiesto en el criterio de normalidad, la comprobación de la *homocedasticidad u homogeneidad de las varianzas* no se ha realizado de forma tan exhaustiva. Aun así, se constata que en más del 50% de los análisis no se puede asumir la igualdad de varianzas entre los grupos comparados.
- El procedimiento de muestreo desarrollado no es probabilístico, de modo que el análisis de la muestra no debería efectuarse por medio de técnicas que partan de tal premisa.

6.1.2 Elección de los procedimientos de análisis no paramétrico.

Definida la postura de partida, para responder a las inferencias estadísticas asociadas a los objetivos E2, G y F2 del marco empírico se ha recurrido a tres tipos de pruebas no paramétricas (ver tabla 89). Considerado el tamaño de la muestra consultada y que los propósitos de la investigación no son de alto riesgo, dichas pruebas se han realizado adoptando un nivel de significación de .05 ($p < .05$). Ello supone asumir en la decisión estadística un margen de error del 5% y un intervalo de confianza (IC) del 95%.

Tabla 89
Pruebas de inferencia estadística no paramétrica empleadas

ANÁLISIS	PRUEBA	EQUIVALENTE PARAMÉTRICO
Comparar los promedios obtenidos en dos grupos.	U-Mann Whitney. Se aplica sobre variables dependientes en escala ordinal o de intervalo; el criterio de agrupación lo determina una variable dicotómica.	<i>t</i> de student
Comparar los promedios obtenidos en tres o más grupos.	H de Kruskal Wallis. Se aplica sobre variables dependientes en escala ordinal o de intervalo; el criterio de agrupación lo determina una variable <i>politómica</i> .	ANOVA
Determinar el tamaño y signo de la correlación lineal entre dos variables.	Rho de Spearman. Se aplica sobre variables cuantitativas continuas ordenadas por rangos. Recurre al coeficiente de correlación r_s para determinar la magnitud de la asociación lineal.	<i>r</i> de Person

Nota: Síntesis elaborada a partir de las contribuciones de varias/os autoras/es (R. Álvarez, 1995; Berlanga & Rubio, 2012; Morales, 2008a; Siegel, 1975).

Para completar las conclusiones derivadas de las pruebas de contraste de medias (*U* de Mann Whitney y *H* de Kruskal Wallis), el análisis realizado incluye el cálculo del tamaño del efecto (TE) (*effect size*). Este complemento aporta la cuantificación de las diferencias evidenciadas, mejorando los resultados obtenidos a través de los métodos tradicionales, centrados en la detección de significación estadística (Iglesias & Sulé, 2005; Morales, 2012; Morse, 1999). La magnitud o tamaño del efecto determina la diferencia media estandarizada entre los grupos comparados (Coe, 2000). Ello permite interpretar los resultados desde su significación práctica, y no sólo estadística, valorando las diferencias y los efectos en función de los propósitos de la investigación y la naturaleza del objeto de estudio (Frías, Pascual, & García, 2000).

El procedimiento aplicado para la estimación del tamaño del efecto ha sido la diferencia tipificada (*d* de Cohen). Para su cálculo, se ha empleado el programa Minsize2 (*Computation of Minimum Sample Size for Statistical Signification*) creado por Morse (1998) para determinar la magnitud del efecto y el tamaño mínimo de muestra. Dicho programa permite calcular el tamaño del efecto en pruebas estadísticas paramétricas y no paramétricas y entre dos o más grupos. Minsize2 informa en eta cuadrada (η^2) la medida de la fortaleza de la relación. Para la clasificación de su magnitud, Morse (1999) convierte a esta métrica las orientaciones clásicas de Cohen (1988; cit. en Morse, 1999) y establece cuatro referencias: $\eta^2 < 0,01$ (efecto muy pequeño), $\eta^2 = 0,01$ (efecto pequeño), $\eta^2 = 0,06$ (moderado) y $\eta^2 = 0,14$ (efecto grande).

Aunque el tamaño del efecto se emplee habitualmente para indagar sobre las diferencias estadísticamente significativas, también puede representar un complemento interesante cuando éstas carezcan de significación, en especial, en el caso de muestras

pequeñas (Morales, 2012). Por razones de viabilidad, esta primera aproximación al estudio de la competencia evaluativa se ha rehusado dicha posibilidad y el cálculo se realiza sólo sobre las diferencias estadísticamente significativas.

Para analizar las correlaciones entre las variables consideradas se ha recurrido a la Rho de Spearman, mejor equivalente no paramétrico a la r de Pearson según R. Álvarez (1995) o Morales (2012). La interpretación de la magnitud de la asociación se ha realizado de acuerdo a las orientaciones de Bisquerra (1987) y Morales (2008a), citadas previamente (tabla 68). El carácter idiográfico y formativo de la presente investigación justifica la elección de estos criterios frente a otras propuestas más exigentes para la interpretación de las asociaciones, como la de Mateo (2009).

6.2 Análisis cualitativos previstos para la consecución de los objetivos.

Entre los propósitos declarados para la investigación en curso (tabla 57) hay tres que requieren una aproximación cualitativa, alejada de los procedimientos cuantitativos presentados en el anterior apartado. Los objetivos E3 y F3 pretenden identificar las principales fortalezas y debilidades para el ejercicio de la evaluación del aprendizaje desde la perspectiva docente y discente y el objetivo H se orienta a la descripción de las experiencias y demandas formativas del profesorado en este terreno.

Con menor relevancia e incidencia que los análisis cuantitativos descritos en el apartado anterior, el estudio cualitativo se ha llevado a cabo sobre las cuestiones abiertas o semi-abiertas incluidas en las herramientas de recogida empleadas en ambas muestras. Dichas preguntas se plantean como complemento del análisis cuantitativo y su misión es contrarrestar la rigidez y superficialidad de las respuestas obtenidas por esa vía.

Por su carácter subsidiario, este proceso se ha realizado sin recurrir a programas específicos para el análisis cualitativo⁸⁹ como Atlas.ti o Ethnograph. Las respuestas registradas en ambas secciones de la muestra han sido concretas y reiteras, estrechamente ligadas a las cuestiones tratadas en los cuestionarios; en consecuencia, su categorización no ha requerido la aplicación de este tipo de herramientas.

⁸⁹ También conocidos como AQDAS-Computer-Aided Qualitative Data Analysis.

7. El capítulo a grandes rasgos

El séptimo capítulo ha reparado en la presentación y justificación del estudio empírico previsto a partir modelo de competencia evaluativa docente elaborado. La tabla 90 detalla las principales decisiones adoptadas en su configuración y los rasgos resultantes:

Tabla 90
Resumen de los principales rasgos del marco empírico de la investigación

DECISIONES	CONCLUSIONES
DISEÑO	• No experimental. Descriptivo, correlacional y comparativo-causal. • Finalidad descriptiva y explicativa. • Dimensión cronológica descriptiva y sentido prospectivo. • Trabajo de campo desarrollado con relación participativa del investigador. • Proceso inductivo. • Estudio de caso. • Dimensión temporal transversal, descriptiva y con matices exploratorios. • Criterio de datación y enfoque metodológico mixto, con preponderancia de lo cuantitativo frente a lo cualitativo.
POBLACIÓN Y MUESTRA	• Población. Estudiantes y profesorado vinculados a la formación básica de maestras/os de Educación Infantil y Primaria en dos centros: Facultad de Psicología y Educación de la Universidad de Deusto y el centro adscrito BAM (Begoñako Andra Mari). • Procedimiento de muestreo no probabilístico (accidental o por accesibilidad y con ciertas cuotas). • Muestra. 51 representantes del profesorado y 704 estudiantes de los tres primeros cursos de ambos centros y titulaciones (Grado en Educación Infantil y Educación Primaria).
VARIABLES	• El estudio abarca la colección de variables dependientes e independientes recogidas en la tabla 61. Las variables dependientes están estrechamente relacionadas con los indicadores de competencia evaluativa determinados en el modelo teórico de referencia. Detallan su nivel de relevancia y el grado de realización práctica que se les atribuye en la muestra. Las variables independientes representan factores demográficos, institucionales, académicos o profesionales con posible efecto en tal valoración.

DECISIONES	CONCLUSIONES
INSTRUMENTOS	• Cuestionario docente (ANEXO 14). Aplicación autónoma y no presencial. Comprende 19 indicadores de la competencia evaluativa estructurados en seis factores: Planificación estratégica de la evaluación, Regulación del aprendizaje y de la enseñanza por medio de la evaluación, Metaevaluación, Ajustes en la evaluación y participación discente, Adecuación y precisión y Puesta en práctica de la evaluación. • Cuestionario discente (ANEXO 15). Administración presencial y dirigida en horario lectivo. Comprende 20 indicadores de competencia evaluativa docente agrupados en seis factores: Participación del alumnado en la evaluación, Planificación del sistema evaluativo, Ajustes del sistema evaluativo, Regulación del aprendizaje por medio de la evaluación, Adecuación y precisión y Control de las condiciones en las que se desarrolla la evaluación. • Los rasgos psicométricos de ambas herramientas en la muestra son aceptables para los propósitos de la investigación pero están sujetas a sugerencias de mejora inicialmente plasmadas en el ANEXO 20.
RECOGIDA DE INFORMACIÓN	• Dos fases de recogida de información para acceder a alumnado y profesorado de ambas titulaciones y centros: segundo cuatrimestre del curso 2011/2012 y primer cuatrimestre de 2012/2013. • La recopilación de las valoraciones discentes se ha realizado en sesiones presenciales dirigidas y las docentes por medio de la cumplimentación autónoma de cuestionarios.
ANÁLISIS PREVISTO	• Análisis descriptivos univariable y análisis inferenciales con pruebas de decisión estadística no paramétrica. por ostensible incumplimiento de los supuestos paramétricos. Para el contraste de medias se establecen como procedimientos de referencia la <i>U</i> de Mann Whitney y la <i>H</i> de Kruskal Wallis según pretendan compararse los promedios de dos o más grupos. Para los estudios de correlación se recurre a la Rho de Spearman. • Intervalo de confianza del 95% y nivel de significación del 5% ($p < .05$). • Cálculo del tamaño del efecto en las diferencias estadísticamente significativas halladas. • Análisis cualitativo complementario.

El próximo capítulo detalla los resultados a los que ha dado lugar el desarrollo de la aproximación empírica a la competencia evaluativa docente, tal y como ha quedado planteada.

CAPÍTULO 8.

PRESENTACIÓN DE RESULTADOS DEL MARCO EMPÍRICO DE LA INVESTIGACIÓN

Las próximas páginas recogen los principales resultados del estudio empírico sobre la competencia evaluativa presentado en el capítulo anterior. Su desarrollo tiene lugar en seis apartados principales:

- A fin de contextualizar los resultados expuestos a lo largo del capítulo, el primer apartado realiza una descripción de la muestra docente y discente a partir de las variables socio-demográficas y académico-profesionales presentes en las mismas.
- Los cuatro siguientes apartados abordan respectivamente los resultados obtenidos en los cuatro objetivos asociados al marco empírico. En cada uno de ellos, la exposición se ordena en dos niveles. En primer lugar, se plantea la lectura o interpretación de los resultados obtenidos en los análisis practicados. A partir de esos datos, en segundo lugar se presenta un breve extracto de las principales conclusiones que plantea su revisión y un avance de la discusión entablada para las mismas en el próximo capítulo.
- El apartado final ofrece una síntesis de las principales conclusiones comunicadas en el capítulo, nuevamente ordenada por objetivo.

Como ya ha sido justificado en el capítulo anterior, la interpretación de los resultados presentados debe partir de dos reservas: la primera hace referencia a la realización del factor sexto de la muestra docente, que sólo será considerado a efectos descriptivos dada su baja consistencia interna, y la segunda se produce en la valoración de los resultados procedentes del factor sexto del alumnado, que ofrece ciertos indicios de ambigüedad que llaman a la cautela.

1. Descripción de la muestra

El diagnóstico de la competencia evaluativa docente que plantea la presente investigación se ha realizado contando con la participación de una muestra de 704 estudiantes y 51 docentes. Expuestos en el capítulo anterior los aspectos metodológicos ligados a su tamaño y representatividad (centro, titulación y curso), este apartado plantea la descripción de cada sector en términos socio-demográficos y académico-profesionales.

La **muestra docente** está conformada por 13 hombres y 38 mujeres con una edad media próxima a los 44 años ($M = 43,76$). El 83,3% de la misma se dedica a la docencia universitaria de forma exclusiva, frente a un porcentaje minoritario de docentes (el 16,7%) que desempeña funciones en otras instituciones educativas fuera de la universidad (fundamentalmente enseñanza o asesoramientos en centros educativos de Educación Básica y/o Secundaria). En cuanto a los tipos de dedicación laboral, más de la mitad del profesorado consultado (el 54,2%) tiene un contrato de 30 horas semanales (dedicación plena); las dedicaciones exclusivas (37,5 horas semanales) representan 29,2% de la muestra y las parciales (menos de 30 horas semanales) el 16,7% restante.

La experiencia docente acumulada resulta muy variable dentro la muestra, oscilando entre los 6 meses y los 32 años. El promedio de dedicación a la enseñanza universitaria es de 12,45 años, pero se constata un alto grado de dispersión en las respuestas ($s = 9,01$). Como resume la figura 34, el profesorado se distribuye en tres grupos de experiencia:

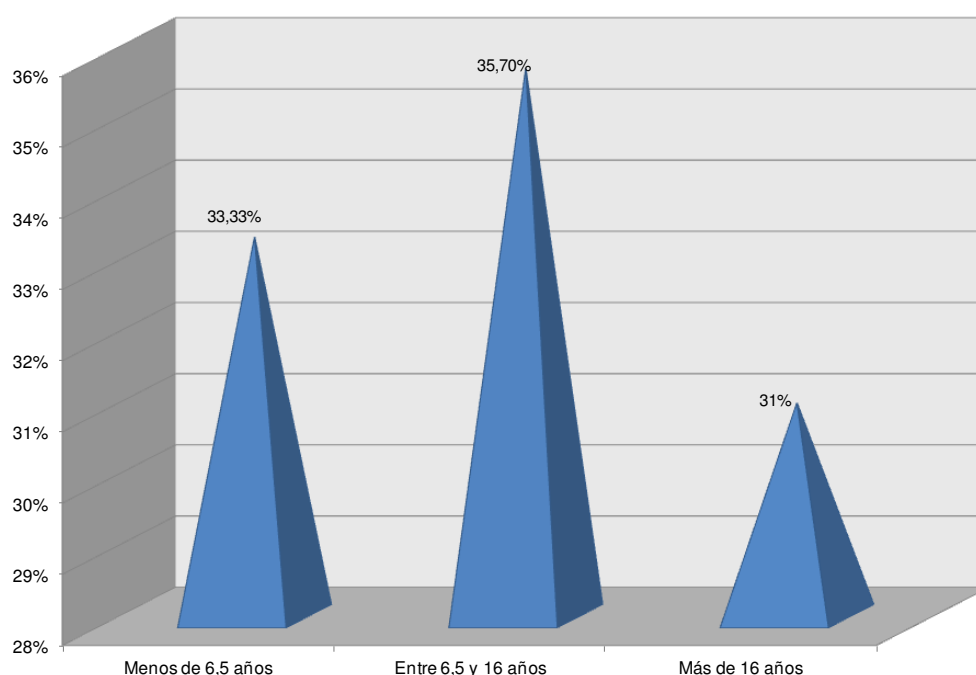


Figura 34. Distribución de la muestra docente por años de experiencia en la enseñanza universitaria.

La formación profesional previa del profesorado también es variada dentro de la muestra (ver tabla 91). Los grupos profesionales más numerosos los conforman las/os licenciadas/os en Pedagogía (33,3%), en Filología (25%) y en Psicología (14,6%).

Tabla 91

Distribución de la muestra docente por titulación de acceso (formación previa)

Formación Previa	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
Pedagogía	16	31,4	33,3	33,3
Filosofía	3	5,9	6,3	39,6
Psicología	7	13,7	14,6	54,2
Psicopedagogía	2	3,9	4,2	58,3
Filología	12	23,5	25,0	83,3
Sociología	2	3,9	4,2	87,5
Actividad Física y Deporte (AFD)	1	2,0	2,1	89,6
Danza	1	2,0	2,1	91,7
Teología	2	3,9	4,2	95,8
Ciencias (Biología)	2	3,9	4,2	100,0
Total válido	48	94,1	100,0	
Sin respuesta (perdidos)	3	5,9		
Total	51	100		

La formación de postgrado dentro de la muestra tiene lugar por medio de estudios al nivel de máster y/o doctorado (ver figura 35). Cerca del 43% del profesorado participante declara haber finalizado algún máster y un 6,10% está en proceso de hacerlo en el momento de la recogida. La proporción de doctoras/es en la muestra es del 26,5% y existe un porcentaje importante de docentes inmersas/os en proceso de tesis doctoral.

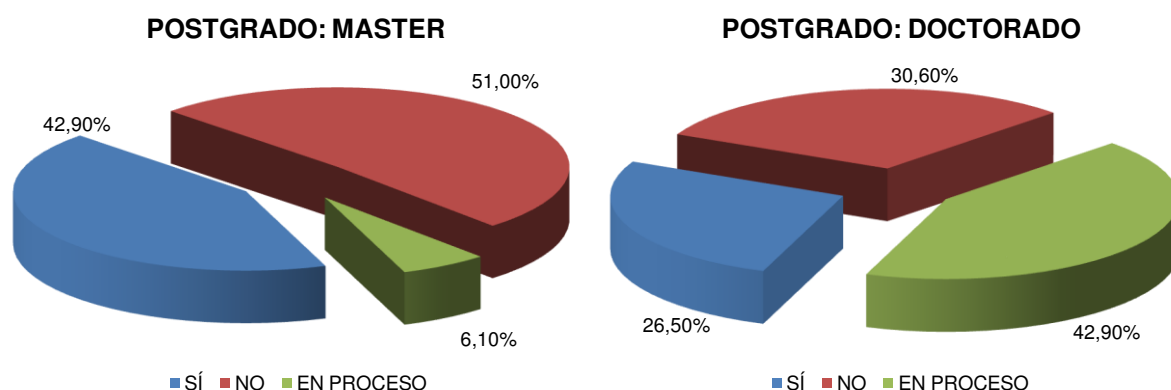


Figura 35. Formación de postgrado en la muestra docente.

El diagnóstico de la competencia evaluativa se ha llevado a cabo tomando como referencia una determinada unidad de docencia. La mayor parte de las valoraciones se ha desarrollado en el contexto de asignaturas completas de 6 ECTS (84,3%) y sólo en 8 casos se ha valorado el sistema evaluativo aplicado en un módulo particular (3 ECTS). En total, la recogida de datos ha tenido lugar en 32 unidades de docencia diferentes, vinculadas a cuatro tipos de módulo dentro de las titulaciones de referencia: formación básica (el 43,1%), didáctico disciplinar (39,2%), mención (13,7%) y Declaración Eclesiástica de Competencia Académica (DECA) (3,9%).

La **muestra discente** está conformada mayoritariamente por mujeres, de modo que los alumnos representan sólo el 35% de los casos. La edad media de las/os participantes se sitúa levemente por debajo de los 20 años ($M = 19,7$; $s = 2,10$), siendo el mínimo 17 años y 41 el máximo. Según refleja la figura 36, el alumnado de menor edad conforma el grupo más numeroso dentro de la muestra.

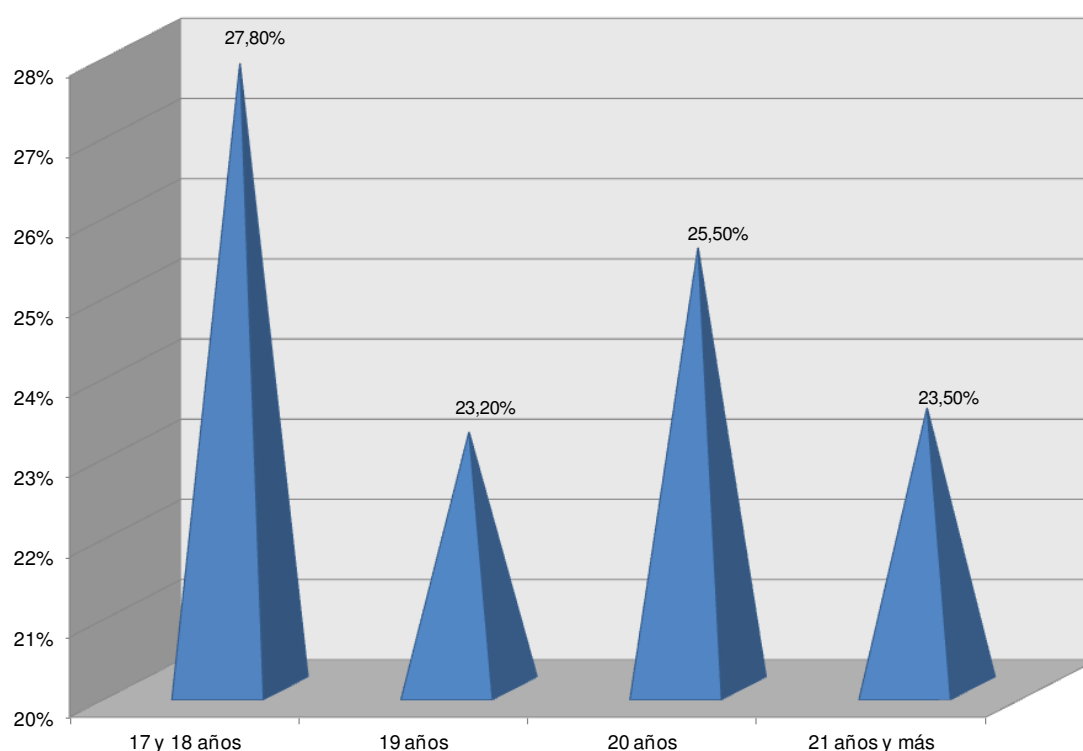


Figura 36. Distribución de la muestra de estudiantes por grupos de edad.

La vía de acceso prioritaria de las/os estudiantes al Grado de Educación Infantil o Primaria es el Bachillerato y la subsiguiente prueba de acceso (86,2%). Un 6,3% del alumnado se inicia en la formación básica de maestras/os habiendo realizado previamente un módulo de Formación Profesional (FP) y otro 5,6% tras haber abandonado otros estudios

universitarios. Las/os estudiantes accedidos por las pruebas para mayores de 25 años o con una titulación universitaria previa representan una clara minoría (el 0,9% y 1,1%, respectivamente).

Como refleja la figura 37, la mayor parte del alumnado tiene dedicación exclusiva a su formación universitaria y no desempeña ninguna actividad remunerada fuera de la misma. Sólo el 8,5% de la muestra desempeña alguna actividad laboral regulada contractualmente.

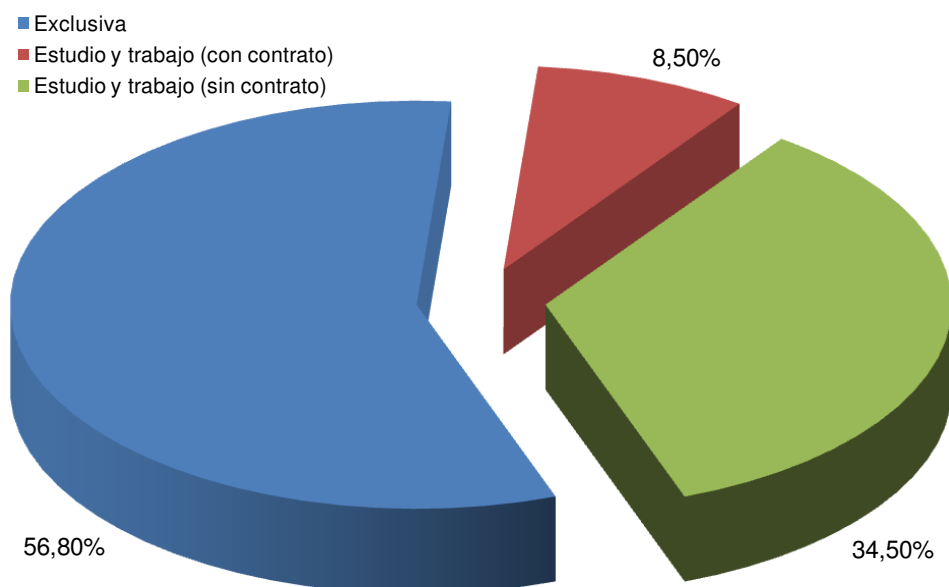


Figura 37. Dedicación de la muestra de estudiantes a su propia formación universitaria.

2. Resultados relativos al Objetivo E. La competencia evaluativa docente desde la perspectiva del profesorado

El primer objetivo del marco empírico hace referencia a la autoevaluación que realiza la muestra docente de su propia competencia evaluativa, atendiendo a la importancia atribuida y al nivel de realización percibido en los 19 indicadores cuestionados y en los seis factores que los agrupan. Incluye tres sub propósitos, todos ellos referidos a la visión del profesorado: el primero tiene carácter descriptivo, el segundo se orienta a la realización de inferencias estadísticas bivariable con respecto a la población de referencia y el tercero

refiere a las valoraciones cualitativas efectuadas sobre los apoyos y las dificultades a la hora de evaluar el aprendizaje del alumnado.

2.1 Resultados Objetivo E1. Identificar los factores e indicadores de la evaluación del aprendizaje en los que mayor y menor dificultad expresa el profesorado, atendiendo al nivel de relevancia y realización que les atribuyen.

El punto de partida en este primer objetivo es que la amplia mayoría del profesorado consultado se considera competente evaluando el aprendizaje de las/os estudiantes (el 84,8%). El porcentaje de la muestra que se declara muy competente alcanza sólo el 2,2% y el 13% restante se autodefine como poco competente en este espacio.

El promedio global de respuesta en la muestra docente sitúa la competencia evaluativa como un aspecto *relevante* ($M = 4,01$), llevado a cabo por debajo de la media teórica ($M = 2,93$, sin llegar a 3 que denota un *cumplimiento adecuado*). Medidas en una escala de intervalo discreta, la dispersión registrada en ambas valoraciones resulta leve ($s = 0,5$ y $s = 0,51$, respectivamente), lo que denota un nivel de acuerdo razonable entre las/os participantes. Trasladados estos resultados a la rejilla AIR (Análisis de Importancia–Realización) de representación (Martilla & James, 1977), la competencia evaluativa docente autopercibida se ubica en el área de Concentración, reservada a aspectos importantes, en los que no se alcanza un nivel de cumplimiento adecuado. El **ANEXO 23** expone brevemente en qué consiste dicha técnica así como la redefinición planteada sobre la misma en la investigación en curso.

El análisis de la competencia evaluativa docente se ha llevado a cabo respetando la estructura de seis factores derivada de las respuestas obtenidas de la muestra. En el contraste de los resultados obtenidos en cada uno de ellos (ver tabla 92), la Regulación del aprendizaje y la enseñanza por medio de la evaluación (F2) se desmarca como el factor más valorado en relevancia ($M = 4,42$) y con mayor grado de acuerdo ($s = 0,51$, la menor de todas). Le sigue de cerca el factor primero sobre la Planificación estratégica de la evaluación del aprendizaje, a la que el profesorado atribuye un nivel de importancia sensiblemente inferior ($M = 4,25$) también con amplia coincidencia ($s = 0,58$). Ambos factores se sitúan por encima del promedio global de importancia calculado para la muestra, denotando una evidente prioridad desde la perspectiva teórica.

En el extremo opuesto, el factor cuarto, asociado a los Ajustes en evaluación y la participación discente, figura como el menos relevante de los considerados ($M = 3,65$, entre *moderadamente importante* e *importante*). La Metaevaluación (F3) y la Adecuación y fiabilidad de la evaluación (F5) obtienen promedios algo más próximos al 4 (*relevante*), pero su nivel de dispersión refleja desacuerdos considerables en su estimación ($s = 0,92$ y $s = 0,86$, respectivamente).

Tabla 92
Valoración de la muestra docente en los seis factores asociados a la competencia evaluativa

Estadísticos descriptivos					
Factores (F)	N	Mínimo	Máximo	Media	s
F1. Planificación evaluación. IMP	51	2,50	5,00	4,2598	0,58301
F1. Planificación evaluación. REA	50	1,75	5,00	3,3450	0,71230
F2. Regulación aprendizaje y enseñanza. IMP	51	2,75	5,00	4,4216	0,51597
F2. Regulación A y E ^a . REA	50	2,25	4,75	3,2250	0,56975
F3. Metaevaluación. IMP	50	1,67	5,00	3,7800	0,92976
F3. Metaevaluación. REA	49	1,00	5,00	2,3878	0,83707
F4. Ajustes y participación. IMP	51	1,75	5,00	3,6503	0,73937
F4. Ajustes y participación. REA	50	1,25	4,75	2,6150	0,74951
F5. Adecuación y precisión. IMP	51	1,50	5,00	3,7451	0,86817
F5. Adecuación y precisión. REA	50	1,00	5,00	2,6000	0,85714
F6. Puesta en práctica. IMP	51	2,00	5,00	4,0000	0,65574
F6. Puesta en práctica. REA	50	1,50	5,00	3,2700	0,70138
N válido (según lista)		51			

Nota: La fuente azul destaca los factores con mayor promedio de puntuación en las subescalas de importancia (IMP.) y realización (REA.). La fuente roja señala los menores promedios registrados en ambos casos.

^a Regulación del aprendizaje y la enseñanza (A y E).

Atendiendo a las valoraciones de realización autopercebida del profesorado, la Planificación estratégica de la evaluación (F1) representa el factor más presente en la práctica ($M = 3,34$, *realización adecuada*). Con promedios sensiblemente inferiores ($M = 3,27$ y $M = 3,22$, respectivamente) le siguen los factores sexto (Puesta en práctica) y segundo (Regulación del aprendizaje y la enseñanza por medio de la evaluación). Los tres factores restantes se sitúan por debajo de la media teórica (3, *cumplimiento adecuado*) y del promedio de realización global detallado para la muestra. La Metaevaluación (F3) es el factor menos trasladado a la práctica, con un cumplimiento *poco adecuado* ($M = 2,38$). Como en la subescala de importancia, la desviación típica asociada al mismo es considerable ($s = 0,837$) y denota claras diferencias en las valoraciones registradas.

Representados en la rejilla AIR (ver figura 38), los seis factores asociados a la competencia evaluativa ocupan el área de Concentración, caracterizada por el desajuste entre la relevancia teórica y el grado de puesta en práctica. La Planificación estratégica de la evaluación (F1) es el que más se aproxima a la zona de Mantenimiento por concentrar el mayor promedio de realización autopercebida; aunque su cumplimiento supera lo *adecuado* en opinión del profesorado, su desarrollo práctico no resulta suficiente para desmarcarse del área de mejora⁹⁰. De modo similar, los factores segundo (Regulación del aprendizaje y la enseñanza por medio de la evaluación) y sexto (Puesta en práctica de la evaluación) también se quedan a menos de medio punto del eje de corte entre los dos cuadrantes superiores. En correspondencia con los resultados revelados, el mayor desajuste se produce en la Metaevaluación (F3), donde la diferencia de promedios entre la importancia reconocida y el cumplimiento declarado dista en 1,39 puntos (tabla 92).

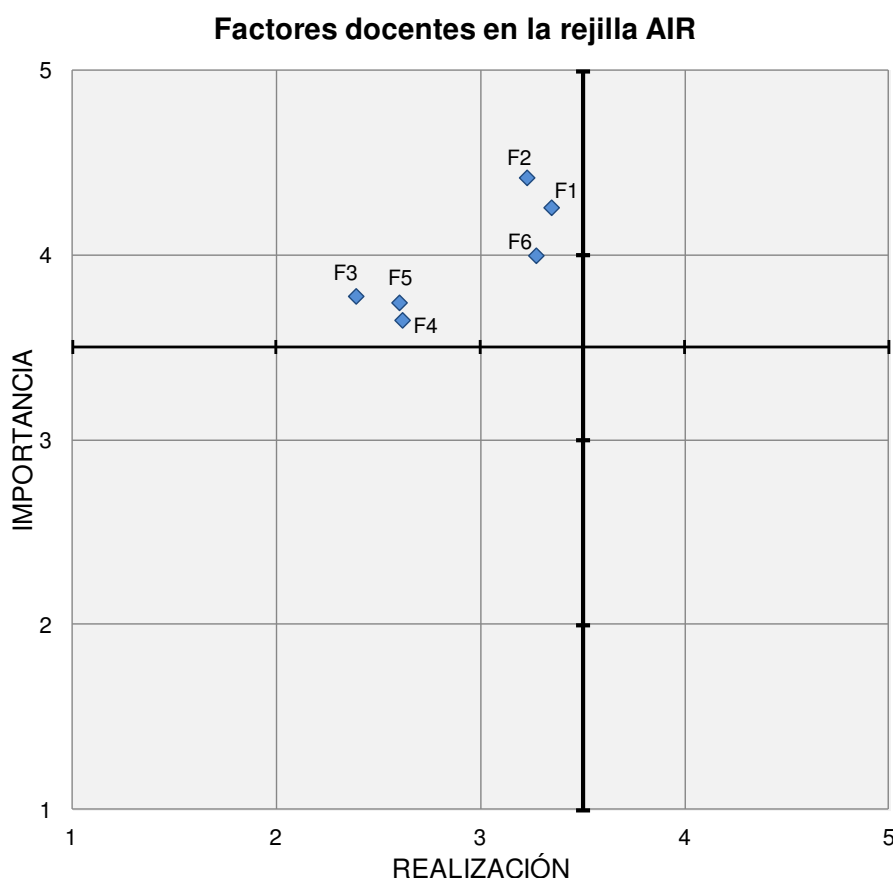


Figura 38. Representación AIR de los factores asociados a la competencia evaluativa docente en la muestra de profesoras/es.

⁹⁰ Según se expone en el **ANEXO 23**, para ello se ha fijado un nivel mínimo de realización autopercebida de 3,5 (*entre adecuado y muy adecuado*). Este criterio representa una condición autoimpuesta para lograr mayor claridad en la delimitación de las zonas ajustadas y desajustadas de la rejilla AIR.

Los factores se presentan en el mismo orden que han sido definidos: F1. Planificación estratégica de la evaluación; F2. Regulación del aprendizaje y de la enseñanza por medio de la evaluación; F3. Metaevaluación; F4. Ajustes en la evaluación y participación discente; F5. Adecuación y precisión; F6. Puesta en práctica de la evaluación.

Los promedios de importancia y realización registrados en los 19 indicadores de evaluación incluidos en el cuestionario revela resultados consistentes con las conclusiones expuestas hasta el momento (la síntesis representada en la figura 39 se encuentra ampliada en el **ANEXO 24**). Los ítems que mayor promedio de relevancia concentran son los referidos a la *comisión de feedback* y a la *regulación de la enseñanza* a partir de los resultados de aprendizaje manifestados en la evaluación (C.6.1. y P.6.3., respectivamente), ambos integrados en el factor segundo, el más valorado de la subescala las/os docentes. Centrado en la *adecuación de la evaluación a la diversidad*, el ítem C.3.2. destaca por ser el menos importante para la muestra ($M = 3,37$) además del que mayor controversia suscita ($s = 1,09$). Le sigue de cerca el ítem P.4.2. (*documentación de los ajustes en el sistema de evaluación*) con un promedio de importancia de 3,64 y una dispersión cercana al punto ($s = 0,98$). Ambos indicadores forman parte del factor cuarto sobre Ajustes en la evaluación y participación discente, el que menor promedio de relevancia presenta y uno de los tres que menos reconoce realizar el profesorado.

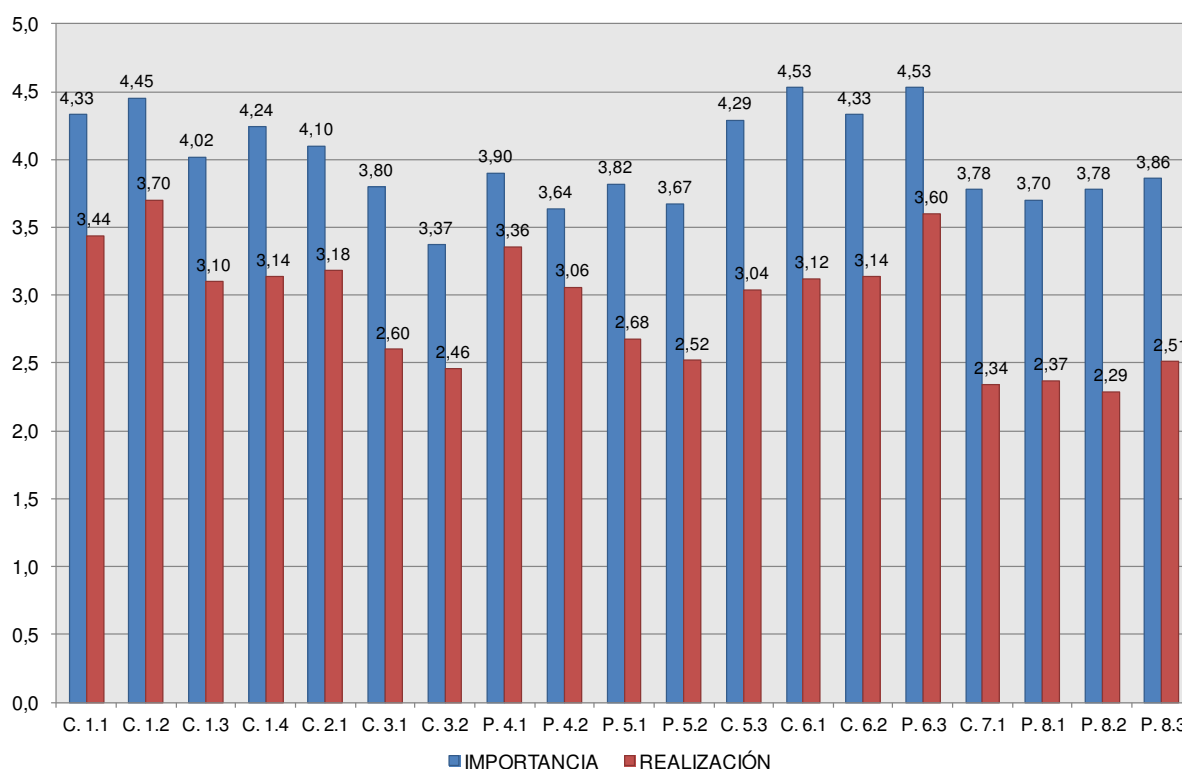


Figura 39. Promedios de importancia y realización declarados por la muestra docente en los 19 indicadores de la competencia evaluativa docente.

De acuerdo a lo expuesto en el ANEXO 17, los ítems encabezados con C son comunes para docentes y alumnado y los precedidos por P, exclusivos para el profesorado.

Del mismo modo, el ítem con menor promedio de realización percibida es el que hace referencia a la *revisión técnica del sistema evaluativo* (P. 8.2.), relacionado con la Metaevaluación (F3), el factor menos valorado en la subescala. La *participación de las/os estudiantes en la evaluación* (C.7.1.) también registra un promedio de realización *poco adecuado* a juicio docente ($M = 2,34$). Por contra, el mayor índice de cumplimiento autopercebido se produce en el ítem C.1.2., que plantea el diseño de *propuestas evaluativas alineadas con los objetivos formativos y la metodología*. Con un promedio de 3,77 (entre *adecuado* y *muy adecuado*) le sigue la *regulación de la enseñanza* en vista de los resultados de aprendizaje del alumnado (P.6.3.). Estos dos ítems están integrados dentro de los dos factores que en mayor medida declara realizar el profesorado de la muestra, a saber, el segundo y el primero.

En comparación con los resultados detallados para los factores, la dispersión registrada en la valoración de los indicadores de competencia evaluativa resulta mucho mayor⁹¹ (**ANEXO 24**). Estas discrepancias resultan especialmente evidentes en la subescala de realización autopercebida, donde siete ítems (el 36,8% del total) muestran índices de desviación típica por encima del punto. La subescala de importancia también presenta una desviación típica media de 0,849.

El traslado de estas valoraciones a la rúbrica empleada para su recopilación ofrece una imagen más descriptiva de la visión del profesorado sobre su propia competencia evaluativa. Su consulta puede realizarse en el **ANEXO 25** como complemento a la interpretación cuantitativa llevada a cabo.

Breve avance de la discusión de resultados...

Coincidiendo con las conclusiones de Villa et al. (2013), en términos generales, el profesorado se califica competente en materia evaluativa. Sin embargo, reconoce dificultades en ciertos factores y manifiesta necesidad de apoyo formativo, igual que en el estudio desarrollado por Tejedor (1999).

En correspondencia con el discurso que se desprende de la literatura específica, la función reguladora que ejerce la evaluación sobre el aprendizaje y la enseñanza se señala como el más

⁹¹ El análisis del siguiente objetivo, profundiza en la naturaleza y significación de tales diferencias, así como en los factores socio-demográficos, profesionales o académico institucionales que parecen mediar en las mismas.

importante de los factores asociados a ella. Como en estudios previos (Arribas et al., 2010; Gutiérrez et al., 2013; L. F. Martínez et al., 2012; G. Rodríguez et al., 2012), la muestra declara para el mismo un nivel de solvencia razonable, aunque en su ejercicio se acusa la incidencia negativa de ciertas condiciones prácticas (objetivo E3). La planificación de la evaluación destaca como el espacio en el que mejor cumplimiento se reconoce para los indicadores de calidad considerados. Este resultado secunda la tendencia de mejora progresiva que se viene señalando los últimos años para algunas de las clásicas deficiencias asociadas al diseño evaluativo (mayor continuidad, transparencia, variedad de herramientas, etc.) (Buscà, Pintor et al., 2010; Ibarra & Rodríguez, 2010; Olmos & Rodríguez, 2010; Padilla et al., 2010).

Confirmando las apreciaciones que se desprenden de la literatura (Monereo & Castelló, 2009a; Monescillo & Toscano, 2007; Santos & Moreno, 2004), la metaevaluación es el factor menos llevado a cabo, a pesar de que el profesorado lo reconozca como relevante. Diferente situación se evidencia para los ajustes en la evaluación y la participación discente, juzgado como el componente menos importante de la competencia evaluativa. Igual que en otras investigaciones (Gutiérrez et al., 2011; G. Rodríguez et al., 2012; Ruiz-Gallardo et al., 2013), el profesorado califica como poco frecuentes las experiencias de autoevaluación y coevaluación. En cuanto a la adecuación o diversificación de las condiciones evaluativas, las evidencias recabadas en ésta y otras muestras sugieren un cierto recelo y desacuerdo que bloquea su puesta en práctica (M. L. Pérez et al., 2000; Tejedor, 1999). La práctica evaluativa deseable parece más asociada a la homogeneidad de condiciones que a su divergencia (V. Álvarez et al., 1999; Serrano, 2010), creencia que también se confirma al consultar el punto de vista discente (objetivo F1) y el contenido de la normativa (Real Decreto 1791/2010).

2.2 Resultados Objetivo E2. Identificar relaciones entre distintas variables académicas, socio-demográficas o profesionales y la competencia evaluativa autopercebida.

La exploración de las respuestas docentes en función de las variables socio-demográficas, académicas y profesionales ha dado lugar a un número importante de conclusiones estadísticamente significativas dentro del intervalo de confianza establecido (IC = 95%) (ver tabla 93). Dichas conclusiones resultan del contraste de promedios de importancia y realización autopercebida en el 58,82% de las 17 variables consideradas dentro de este bloque:

- Entre las variables académicas examinadas, observamos que la unidad de docencia considerada (módulo de 3 ECTS o asignatura completa de 6 ECTS), así como el módulo, la materia y el título de grado al que se adscribe la misma, no provocan

diferencias estadísticamente significativas en las respuestas de la muestra. Asimismo, la desigualdad de promedios que se advierte entre profesorado con distinta dedicación (exclusiva, plena o parcial) y formación de postgrado (máster y/o doctorado) sólo puede ser atribuida al azar o al propio al procedimiento de muestreo ($p \geq .05$).

- En las 10 variables restantes, se advierten resultados generalizables a la población de referencia, con un margen de error inferior al 5%. La magnitud media del efecto estimado para los mismos es grande ($\eta^2 = 0,1523^{92}$).

Tabla 93

Conclusiones estadísticamente significativas ($p < .05$) en la valoración de la competencia evaluativa efectuada en la muestra docente por factor y subescala (Sub.)

FACTOR	Sub.	CONCLUSIÓN
TOTAL	IMP.	• SEXO. Las mujeres mayor promedio (TE MEDIO). • FORMACIÓN PSICOLÓGICA, PEDAGÓGICA o PSICOPEDAGÓGICA. Mayor promedio en el profesorado que la tiene que en el que no (TE MEDIO).
	REA.	• FORMACIÓN PREVIA. Licenciatura en Filología menor promedio que en Filosofía (TE GRANDE). • FORMACIÓN PSICOLÓGICA, PEDAGÓGICA o PSICOPEDAGÓGICA. Mayor promedio en el profesorado que la tiene que en el que no (TE MEDIO). • FORMACIÓN FILOLÓGICA. Menor promedio en el profesorado formado en Filología (TE GRANDE). • FORMACIÓN RECIENTE EVALUACIÓN. Mayor promedio en quienes han realizado formación sobre evaluación los dos últimos cursos (TE MEDIO).
FACTOR 1	IMP.	• FORMACIÓN PSICOLÓGICA, PEDAGÓGICA o PSICOPEDAGÓGICA. Mayor promedio en el profesorado que la tiene que en el que no (TE MEDIO). • SIMULTANEIDAD. Menor promedio en el profesorado con funciones docentes en otras instituciones educativas (TE MEDIO).
	REA.	• FORMACIÓN PREVIA. Licenciatura en Filología menor promedio que en Pedagogía/Psicopedagogía (TE GRANDE). • FORMACIÓN PEDAGÓGICA o PSICOPEDAGÓGICA. Mayor promedio en el profesorado que la tiene que en el que no (TE GRANDE). • FORMACIÓN PSICOLÓGICA, PEDAGÓGICA o PSICOPEDAGÓGICA. Mayor promedio en el profesorado que la tiene que en el que no (TE GRANDE). • FORMACIÓN FILOLÓGICA. Menor promedio en el profesorado formado en Filología (TE GRANDE). • CURSO. Mayor promedio en asignaturas de primer curso (TE MEDIO). • COMPETENCIA EVALUATIVA (dos grupos). Menor promedio en el profesorado que se considera poco competente que en el autocalificado como competente o muy competente (TE MEDIO).

⁹² Promedio de las puntuaciones eta cuadrada de todas las diferencias estadísticamente significativas detectadas en la muestra de profesoras/es.

FACTOR	Sub.	CONCLUSIÓN
FACTOR 2	IMP.	-
	REA.	• EDAD. Correlación positiva y baja. • NIVEL EXPERIENCIA. En el grupo de mayor experiencia los promedios son superiores que en el intermedio (TE MEDIO).
FACTOR 3	IMP.	• SEXO. Las mujeres mayor promedio (TE MEDIO). • NECESIDAD FORMATIVA. Quienes manifiestan sentir necesidad formativa en evaluación mayor promedio (TE MEDIO).
	REA.	• CENTRO. Profesorado de la UD mayor promedio (TE MEDIO).
FACTOR 4	IMP.	• FORMACIÓN PSICOLÓGICA, PEDAGÓGICA o PSICOPEDAGÓGICA. Mayor promedio en el profesorado que la tiene que en el que no (TE MEDIO). • FORMACIÓN RECIENTE EVALUACIÓN. Mayor promedio en quienes han realizado formación sobre evaluación los dos últimos cursos (TE MEDIO).
	REA.	• FORMACIÓN PREVIA. No se producen diferencias en las licenciaturas en Biología, Sociología y Teología, pero sí en todas las restantes (TE GRANDE). • FORMACIÓN PSICOLÓGICA, PEDAGÓGICA o PSICOPEDAGÓGICA. Mayor promedio en el profesorado que la tiene que en el que no (TE GRANDE). • FORMACIÓN FILOLÓGICA. Menor promedio en el profesorado formado en Filología (TE GRANDE). • FORMACIÓN RECIENTE EVALUACIÓN. Mayor promedio en quienes han realizado formación sobre evaluación los dos últimos cursos (TE MEDIO).
FACTOR 5	IMP.	• CURSO. Mayor promedio en asignaturas de primer curso (TE GRANDE).
	REA.	• COMPETENCIA EVALUATIVA (dos grupos). Menor promedio en el profesorado que se considera poco competente que en el autocalificado como competente o muy competente (TE MEDIO).
FACTOR 6	IMP.	• FORMACIÓN PEDAGÓGICA/PSICOPEDAGÓGICA. Mayor promedio en el profesorado que la tiene que en el que no (TE MEDIO). • FORMACIÓN PSICOLÓGICA, PEDAGÓGICA o PSICOPEDAGÓGICA. Mayor promedio en el profesorado que la tiene que en el que no (TE MEDIO). • CENTRO. Profesorado de la UD mayor promedio (TE MEDIO).
	REA.	• ~~FORMACIÓN PSICOLÓGICA, PEDAGÓGICA/PSICOPEDAGÓGICA.~~

Nota: Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Planificación estratégica de la evaluación; F2. Regulación del aprendizaje y de la enseñanza por medio de la evaluación; F3. Metaevaluación; F4. Ajustes en la evaluación y participación discente; F5. Adecuación y precisión; F6. Puesta en práctica de la evaluación. La valoración de la realización ligada a este último factor no se considera por la escasa consistencia interna detectada al analizar su fiabilidad.

En coherencia con las conclusiones presentadas en el sub-propósito E1, la importancia atribuida a la Regulación del aprendizaje y la enseñanza por medio de la evaluación (Factor 2) representa la única valoración que no varía en función de las variables socio-demográficas, profesionales, evaluativas o académico-institucionales consideradas. Este dato evidencia nuevamente que existe un acuerdo generalizado sobre su importancia o pertinencia teórica, con independencia de los rasgos específicos que acompañen al profesorado.

En el resto de factores se detecta una amplia gama de diferencias con incidencia en ambas subescalas de valoración. La realización autopercebida en la Planificación estratégica de la evaluación (Factor 1) destaca como la puntuación que mayor número de diferencias significativas concentra. La Metaevaluación (Factor 3) y la Adecuación y fiabilidad de la evaluación (Factor 5) se sitúan en el extremo opuesto, al dar lugar a un único contraste significativo de promedios.

Como ya ha sido justificado en el capítulo anterior, las características del diseño de la investigación en curso, especialmente la falta de control experimental sobre las variables analizadas, limita el establecimiento de relaciones directas de causalidad en las conclusiones presentadas. Con la cautela a la que obliga tal circunstancia, los próximos apartados ordenan los resultados en función de la naturaleza de las variables en las que han sido constatados:

- Variables socio-demográficas (sexo y edad).
- Variables profesionales (años de experiencia docente, titulación de acceso y simultaneidad de la docencia y otros compromisos laborales en instituciones educativas).
- Variables académico-institucionales (curso y centro).
- Variables independientes más directamente vinculadas a la evaluación del aprendizaje (competencia evaluativa autopercebida, realización de formación permanente sobre evaluación del aprendizaje y necesidad autopercebida de formación sobre el tema).

El recorrido de dichos apartados desvela progresivamente las conclusiones anticipadas en la tabla 93.

2.2.1 Variables socio-demográficas y valoraciones docentes.

Los promedios de cumplimiento autodeclarados para los seis factores analizados no difieren de forma significativa entre los profesores y profesoras de la muestra. Sin embargo, en la subescala de importancia la variable **sexo** da lugar a dos diferencias con una significación por debajo del 5% ($p < .05$) y un tamaño de efecto (TE) medio (ver tabla 94). Como refleja la tabla 95, las mujeres obtienen una mayor puntuación que los hombres en el total de la subescala de importancia, lo que implica que han considerado más relevantes los

indicadores asociados a su competencia evaluativa. Así se constata también al estimar la importancia de la revisión del sistema evaluativo o Metaevaluación.

Tabla 94

Autovaloración de la competencia evaluativa en la muestra docente en función del SEXO

Estadísticos de contraste (U de Mann Whitney)

Factores	U de Mann-Whitney	W de Wilcoxon	Z	Sig. asintót. (bilateral)	TE (η^2)
Total IMPORTANCIA	148,500	239,500	-2,132	,033	0,0909 MEDIO
Total REALIZACIÓN	183,500	886,500	-1,262	,207	
FACTOR 1 IMP.	173,000	264,000	-1,618	,106	
FACTOR 1 REA.	176,000	879,000	-1,438	,150	
FACTOR 2 IMP.	197,500	288,500	-1,088	,277	
FACTOR 2 REA.	202,000	905,000	-,863	,388	
FACTOR 3 IMP.	132,500	223,500	-2,415	,016	0,1166 MEDIO
FACTOR 3 REA.	191,000	857,000	-,985	,325	
FACTOR 4 IMP.	214,500	305,500	-,709	,478	
FACTOR 4 REA.	228,000	931,000	-,279	,780	
FACTOR 5 IMP.	224,500	315,500	-,504	,614	
FACTOR 5 REA.	191,500	894,500	-1,120	,263	
FACTOR 6 IMP.	171,500	262,500	-1,679	,093	
FACTOR 6 REA.	~~161,000~~	~~864,000~~	~~-1,823~~	~~,068~~	

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Planificación estratégica de la evaluación; F2. Regulación del aprendizaje y de la enseñanza por medio de la evaluación; F3. Metaevaluación; F4. Ajustes en la evaluación y participación discente; F5. Adecuación y precisión; F6. Puesta en práctica de la evaluación. La valoración de la realización ligada a este último factor no se considera por la escasa consistencia interna detectada al analizar su fiabilidad.

Tabla 95

Diferencias significativas en las valoraciones de la muestra docente en función del SEXO: estadísticos de grupo

Factores	Sexo	n	M	s	SEM
Total IMPORTANCIA	HOMBRE	13	3,7978	,37223	,10324
	MUJER	38	4,0853	,52642	,08540
F3. Metaevaluación IMP.	HOMBRE	13	3,2308	,91676	,25426
	MUJER	37	3,9730	,86559	,14230

En este bloque de variables también se constata una correlación significativa, positiva y baja, entre la **edad** del profesorado y el grado de realización que declara en el factor segundo, sobre la Regulación del aprendizaje y la enseñanza por medio de la evaluación ($r_s = .372$; $p = .017$). Esta asociación indica que a mayor edad, el profesorado

percibe un mayor cumplimiento de la función reguladora de la evaluación en su propio trabajo. Sin embargo, analizada como variable *politómica*, no se detecta ninguna diferencia estadísticamente significativa en las valoraciones de importancia y realización ofrecidas por el profesorado de distintos grupos de edad (39 años y menores, 48 años y mayores y un grupo intermedio entre los 40 y 47 años).

2.2.2 Variables profesionales y valoraciones docentes.

La formación previa del profesorado, los años de experiencia docente y el desempeño exclusivo o simultáneo de la enseñanza universitaria dan lugar a promedios de valoración significativamente diferentes en todos los factores considerados, a excepción del tercero (Metaevaluación).

La correlación de los años de *experiencia docente universitaria* y las estimaciones de relevancia y realización declaradas no produce ningún resultado significativo al nivel establecido ($p < .05$). Sin embargo, la comparación de las respuestas generadas en los tres grandes grupos de experiencia presentes en la muestra (figura 34) revela que la realización autopercebida en el factor segundo difiere significativamente ($p = .029$) y con una magnitud grande de efecto ($\eta^2 = 0,1412$) (ver tabla 96).

Tabla 96

Autovaloración de la competencia evaluativa en la muestra de profesoras/es en función del NIVEL DE EXPERIENCIA DOCENTE

Estadísticos de contraste (<i>H</i> Kruskal-Wallis)				
Factores	Chi-cuadrado	gl	Sig. asintót.	TE (η^2)
Total IMPORTANCIA	,860	2	,651	
Total REALIZACIÓN	4,716	2	,095	
FACTOR 1 IMP.	,237	2	,888	
FACTOR 1 REA.	3,864	2	,145	
FACTOR 2 IMP.	,142	2	,932	
FACTOR 2 REA.	7,058	2	,029	0,1412 GRANDE
FACTOR 3 IMP.	2,748	2	,253	
FACTOR 3 REA.	3,639	2	,162	
FACTOR 4 IMP.	1,257	2	,533	
FACTOR 4 REA.	,461	2	,794	
FACTOR 5 IMP.	1,939	2	,379	
FACTOR 5 REA.	4,544	2	,103	
FACTOR 6 IMP.	,299	2	,861	

Estadísticos de contraste (*H* Kruskal-Wallis)

Factores	Chi-cuadrado	gl	Sig. asintót.	TE (η^2)
FACTOR 6-REA.	1,050	2	,592	

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Planificación estratégica de la evaluación; F2. Regulación del aprendizaje y de la enseñanza por medio de la evaluación; F3. Metaevaluación; F4. Ajustes en la evaluación y participación discente; F5. Adecuación y precisión; F6. Puesta en práctica de la evaluación. La valoración de la realización ligada a este último factor no se considera por la escasa consistencia interna detectada al analizar su fiabilidad.

Para determinar entre qué grupos de experiencia se producen las diferencias en la realización autopercebida en la Regulación del aprendizaje y la enseñanza por medio de la evaluación (Factor 2), se ha recurrido a la exploración *post hoc* que ofrece la alternativa paramétrica del análisis de varianza (ANOVA) (ver tabla 97). Aplicado el contraste de Scheffé se observa que éstas tienen lugar solamente entre los dos grupos de mayor experiencia. El profesorado con más años de experiencia estima realizar un cumplimiento *adecuado* para esta función evaluativa, más de medio punto por encima del declarado en el grupo intermedio.

Tabla 97

Diferencias significativas en las valoraciones de la muestra de profesoras/es en función del NIVEL DE EXPERIENCIA DOCENTE: estadísticos de grupo y contraste post hoc (Scheffé)

Diferencias significativas en función del NIVEL EXPERIENCIA		<i>n</i>	<i>M</i>	<i>s</i>	Sig. (< 6,5)	Sig. (6,5 y 16)	Sig. (> 16)
REA. FACTOR 2	Menor (< 6,5 años)	14	3,3750	,48783	-	0,056	0,962
	Media (6,5 y 16 años)	15	2,8500	,51582	0,056	-	0,38
	Mayor (> 16 años)	12	3,4375	,70004	0,962	0,38	-

Nota: Las celdas sombreadas señalan los grupos entre los que tiene lugar la diferencia de promedios dentro del intervalo de confianza establecido ($p < .05$).

La **formación profesional previa del profesorado**, entendida como su titulación inicial de acceso a la docencia universitaria, revela múltiples diferencias entre las valoraciones producidas en la muestra docente. Partiendo de la respuesta directa ofrecida por el profesorado participante en los datos de identificación del cuestionario, se ha realizado una primera recodificación de la variable titulación en ocho categorías, dando lugar a los promedios de respuesta resumidos en la tabla 98:

Tabla 98

Autovaloración de la competencia evaluativa en la muestra docente en función de la FORMACIÓN PREVIA DEL PROFESORADO

Lic	TOTAL		FACTOR 1		FACTOR 2		FACTOR 3		FACTOR 4		FACTOR 5		FACTOR 6	
	I	R	I	R	I	R	I	R	I	R	I	R	I	R
1	4,12	3,04	4,48	3,75	4,54	3,23	3,62	2,25	3,76	2,85	3,83	2,32	4,27	3,50
2	3,75	3,61	3,66	3,25	4,08	4,08	3,22	3,55	3,75	3,41	4,00	3,66	3,83	3,83
3	4,29	3,13	4,35	3,42	4,57	3,25	4,38	2,85	4,10	2,92	4,07	2,64	4,07	3,64
4	3,98	2,60	4,16	2,81	4,54	3,20	4,02	2,02	3,38	2,08	3,54	2,29	4,04	3,16
5	3,86	2,50	3,87	2,75	4,62	2,75	3,33	1,50	3,50	2,37	4,75	3,25	3,00	2,50
6	3,39	2,55	4,62	3,50	4,00	2,87	2,83	2,50	2,37	1,75	2,75	2,00	3,25	2,25
7	3,60	3,18	4,12	4,00	3,37	3,25	3,66	2,50	3,50	2,75	3,00	3,75	3,75	2,75
8	4,05	3,18	4,37	3,75	4,25	3,12	3,83	3,16	4,00	2,50	4,00	3,50	3,50	3,25
T	4,02	2,94	4,29	3,38	4,43	3,24	3,78	2,39	3,64	2,61	3,77	2,58	4,01	3,31

Nota: La columna de la izquierda (Lic.) recoge la recodificación de las titulaciones profesionales detalladas al describir la muestra docente. Los valores resultantes hacen referencia a las siguientes licenciaturas: 1. Pedagogía o Psicopedagogía; 2. Filosofía; 3. Psicología; 4. Filología; 5. Sociología; 6. Danza o AFD; 7. Teología; 8. Ciencias (Biología); T. Total. Las celdas sombreadas anticipan los factores y valoraciones en las que se producen diferencias significativas ($p < .05$) entre los grupos profesionales.

Comparando las respuestas de los ocho grupos profesionales por medio de la H de Kruskal Wallis se evidencian diferencias significativas de magnitud grande ($\eta^2 > 0,14$) en tres valoraciones de la subescala de realización: la total y las referidas a los factores primero (Planificación estratégica de la evaluación) y cuarto (Ajustes en la evaluación y participación discente) (**ANEXO 26**, fase I). La réplica de este contraste por la vía paramétrica (ANOVA) ha permitido acotar entre qué grupos se producen las diferencias y detectar la especial presencia del profesorado titulado en Pedagogía o Psicopedagogía (grupo 1) y en Filología (grupo 4) en las mismas:

- El contraste de Bonferroni revela que el promedio total de realización autopercebida difiere de forma significativa ($p = .039$) entre el profesorado licenciado en Filosofía y Filología, siendo este último grupo el que menor puntuación presenta (ver tabla 98).
- En el primer factor el contraste de medias significativo se produce entre los grupos profesionales primero y cuarto, a saber, aquel con formación profesional pedagógica o psicopedagógica y el licenciado en Filología ($p = .005$, aplicado en contraste de Bonferroni y $p = .034$ en el de Scheffé). También en este caso el menor promedio de cumplimiento se produce entre las/os filólogas/os (ver tabla 98).
- En la realización autopercebida para el factor cuarto, sobre la Participación discente y los ajustes de la evaluación, las diferencias significativas afectan a cinco de los ocho

grupos profesionales (no se producen diferencias significativas con respecto a las valoraciones de docentes licenciadas/os en Teología, Biología y Sociología).

Para indagar en estas primeras conclusiones, se ha llevado a cabo una segunda fase de análisis, comparando las valoraciones procedentes del profesorado con formación profesional previa pedagógica o psicopedagógica (grupo 1) y las del resto de la muestra (**ANEXO 26**, fase II). Aplicada sobre esta nueva variable dicotómica (con o sin formación pedagógica o psicopedagógica), la *U* de Mann Whitney ha revelado nuevas diferencias dentro del intervalo de confianza establecido. Tanto al valorar la relevancia del factor sexto (Puesta en práctica de la evaluación), como el grado de cumplimiento en la Planificación estratégica de la evaluación (F1), las/os docentes con formación pedagógica o psicopedagógica previa muestran promedios significativamente superiores que el resto de profesionales ($p = .022$; $\eta^2 = 0,1056$ y $p = .002$; $\eta^2 = 0,1829$, respectivamente).

Para seguir profundizando en las diferencias anteriores, en una tercera conversión se han diferenciado en la muestra docente dos grupos profesionales: aquellos con formación profesional previa en Psicología, Pedagogía o Psicopedagogía (grupos profesionales 1 y 3) y el resto de profesionales docentes. Según precisa el **ANEXO 26** (fase III), el contraste de los promedios evidenciados en ambos grupos da lugar a diferencias significativas dentro del intervalo de confianza establecido (IC = 95%; $p < .05$). Éstas se producen en las valoraciones totales de ambas subescalas, así como en la importancia y realización declaradas en los Factores 1, 4 y 6. Con un tamaño de efecto medio para las valoraciones de relevancia (puntuaciones eta cuadrado entre 0,06 y 0,14) y grande para las de realización ($\eta^2 > 0,14$), el denominador común en todas las diferencias es que el profesorado con formación profesional previa psicológica, pedagógica o psicopedagógica presenta puntuaciones superiores que el resto de la muestra.

La cantidad de diferencias evidenciadas en función de la titulación inicial del profesorado y la magnitud del efecto asociado a varias de ellas la señalan como una variable relevante en la competencia evaluativa del profesorado. Por ese motivo, para completar el análisis se han desarrollado cuatro fases más de exploración, además las tres que ya han sido presentadas (ver tabla 99). Como recoge el **ANEXO 26** (fase IV), el siguiente paso ha centrado la atención en el profesorado con formación previa filológica (grupo 4), otro de los más presentes en las diferencias evidenciadas hasta el momento. Comparadas sus respuestas con las del resto de la muestra, se aprecian diferencias estadísticamente significativas ($p < .05$) en la realización autopercebida para el Factor 1, 4 y en el total. Los promedios inferiores se concentran en el grupo de filólogas/os, siendo

grande el tamaño de efecto estimado en los tres casos (las puntuaciones eta cuadrada oscilan entre 0,1524 y 0,2140).

Los resultados de los análisis relatados hasta el momento y resumidos en la tabla 99 apuntan en una misma dirección. Las diferencias halladas en las sucesivas rondas de comparación parecen especialmente evidentes entre dos grupos profesionales en particular (profesorado con formación previa en Psicología, Pedagogía o Psicopedagogía y docentes licenciadas/os en Filología) y en tres valoraciones de la subescala de realización, en las que se constatan efectos de tamaño grande (total y Factores 1 y 4). Las fases quinta y sexta de exploración así lo han verificado, al revelar nuevamente menores promedios de realización en el grupo de tituladas/os en Filología (**ANEXO 26**, fases V y VI).

Tabla 99

Formación previa del profesorado y autoevaluación de su competencia evaluativa

GRUPOS PROFESIONALES	SÍNTESIS DE RESULTADOS
FASE I 8 grupos profesionales (recodificación inicial)	- Realización total menor en Filología que en Filosofía (TE GRANDE). - Realización de Planificación estratégica (F1) menor en filólogas/os que en tituladas/os en Pedagogía o Psicopedagogía (TE GRANDE). - Realización de Ajustes en la evaluación y participación discente (F4) entre diversos grupos profesionales, con especial presencia de pedagogas/os y psicopedagogas/os (TE GRANDE).
FASE II 2 grupos (Pedagogía o Psicopedagogía y resto)	- Realización mayor en Planificación estratégica (F1) (TE GRANDE). - Importancia mayor en Puesta en práctica (F6) (TE MEDIO).
FASE III 2 grupos (Psicología, Pedagogía o Psicopedagogía y resto)	- Importancia y realización total mayor (efecto MEDIO). - Importancia y realización mayor en Planificación estratégica (F1) (TE GRANDE en realización y MEDIO en importancia). - Importancia y realización mayor en Ajustes y participación (F4) (TE GRANDE en realización y MEDIO en importancia). - Importancia mayor en Puesta en práctica (F6) (TE MEDIO).
FASE IV 2 grupos (Filología y resto)	- Realización total menor (TE GRANDE). - Realización menor en Planificación estratégica (F1) (TE GRANDE). - Realización menor en Ajustes y participación (F4) (TE GRANDE).
FASE V 2 grupos (Pedagogía o Psicopedagogía y Filología)	Realización total menor en Filología (TE GRANDE). Realización de Planificación estratégica menor en Filología (TE GRANDE). Realización de Ajustes en la evaluación y participación discente menor en Filología (TE GRANDE).
FASE VI 2 grupos (Psicología, Pedagogía o Psicopedagogía y Filología)	
FASE VII 2 grupos (CC. Naturales y Sociales)	No se producen diferencias significativas.

Nota: Los filetes enmarcan las conclusiones más claras de las fases sucesivas de exploración.

En vista de estas primeras conclusiones, sería interesante ampliar la exploración a otras familias profesionales que no se encuentran suficientemente representadas en la muestra (Filosofía, Ciencias, etc.). El último análisis realizado en esta variable, evidencia que las valoraciones de las/os docentes con formación previa en ciencias sociales no difieren de forma significativa de las que llevan a cabo el profesorado formado en ciencias naturales, pero el número de sujetos incluido en este grupo es muy reducido (**ANEXO 26**, fase VII). Como se expone en el próximo capítulo, la formación previa del profesorado también ha sido considerada en otros estudios relativos a la evaluación del aprendizaje, aunque las conclusiones comunicadas no son determinantes ni consistentes.

La última variable contemplada en este bloque de factores profesionales hace referencia a la **simultaneidad de la docencia universitaria** con otro trabajo en instituciones educativas universitarias o no universitarias. Como ya ha sido indicado en la descripción de la muestra, este trabajo simultáneo se produce en algo menos del 17% del profesorado de la muestra, pero parece asociado a una diferencia de promedios generalizable a la población de referencia (ver tabla 100). Con un efecto de tamaño medio y una significación dentro del intervalo de confianza establecido ($p = .023$; $\eta^2 = 0,1027$), el profesorado que ejerce funciones educativas fuera de la universidad concede menor relevancia a la Planificación estratégica de la evaluación (ver tabla 101).

Tabla 100

Autovaloración de la competencia evaluativa en la muestra docente en función de la SIMULTANEIDAD o EXCLUSIVIDAD en el ejercicio de la enseñanza

Estadísticos de contraste (*U* de Mann Whitney)

Factores	<i>U</i> de Mann-Whitney	<i>W</i> de Wilcoxon	<i>Z</i>	Sig. asintót. (bilateral)	TE (η^2)
Total IMPORTANCIA	120,000	156,000	-1,108	,268	
Total REALIZACIÓN	145,000	925,000	-,312	,755	
FACTOR 1 IMP.	79,000	115,000	-2,266	,023	0,1027 MEDIO
FACTOR 1 REA.	144,000	924,000	-,342	,732	
FACTOR 2 IMP.	101,500	137,500	-1,644	,100	
FACTOR 2 REA.	128,000	908,000	-,801	,423	
FACTOR 3 IMP.	119,500	155,500	-1,133	,257	
FACTOR 3 REA.	118,500	898,500	-1,073	,283	
FACTOR 4 IMP.	142,000	962,000	-,502	,615	
FACTOR 4 REA.	154,500	190,500	-,043	,966	
FACTOR 5 IMP.	146,500	966,500	-,386	,700	
FACTOR 5 REA.	122,000	902,000	-,998	,318	

Estadísticos de contraste (*U* de Mann Whitney)

Factores	<i>U</i> de Mann-Whitney	<i>W</i> de Wilcoxon	<i>Z</i>	Sig. asintót. (bilateral)	TE (η^2)
FACTOR 6 IMP.	134,000	170,000	-,742	,458	
FACTOR 6 REA.	138,500	918,500	-,516	,606	

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Planificación estratégica de la evaluación; F2. Regulación del aprendizaje y de la enseñanza por medio de la evaluación; F3. Metaevaluación; F4. Ajustes en la evaluación y participación discente; F5. Adecuación y precisión; F6. Puesta en práctica de la evaluación. La valoración de la realización ligada a este último factor no se considera por la escasa consistencia interna detectada al analizar su fiabilidad.

Tabla 101

Diferencias significativas en las valoraciones de la muestra docente en función de la SIMULTANEIDAD (SI) o EXCLUSIVIDAD (NO) en el ejercicio de la enseñanza: estadísticos de grupo

Factores	Simultaneidad	<i>n</i>	<i>M</i>	<i>s</i>	<i>SEM</i>
F1. Planificación IMP.	SI	8	3,8125	,60872	,21521
	NO	40	4,3313	,55293	,08743

El resto de las variables de tipo profesional (formación de postgrado a nivel de máster y doctorado, tipo de dedicación o contrato docente) no han dado lugar a diferencias estadísticamente significativas en la muestra de profesoras/es.

2.2.3 Variables académico-institucionales y valoraciones docentes.

Entre las seis variables académico-institucionales consideradas en la muestra docente, los análisis practicados revelan promedios significativamente diferentes en función del **centro** y del curso.

Como refleja la tabla 102, el profesorado de la UD y del centro adscrito BAM realiza valoraciones significativamente diferentes en dos casos: al declarar niveles de realización en el factor Metaevaluación ($p = .031$) y al estimar el grado de importancia de la Puesta en práctica de la evaluación ($p = .024$). Con una magnitud de efecto media en ambos casos, la comparación de los promedios de respuesta revela que es el personal de la UD quien mayores puntuaciones registra en ambos casos (ver tabla 103).

Tabla 102

Autovaloración de la competencia evaluativa en la muestra docente en función del CENTRO

Estadísticos de contraste (U de Mann Whitney)

Factores	U de Mann-Whitney	W de Wilcoxon	Z	Sig. asintót. (bilateral)	TE (η^2)
Total IMPORTANCIA	205,500	800,500	-1,671	,095	
Total REALIZACIÓN	195,000	790,000	-1,603	,109	
FACTOR 1 IMP.	234,500	829,500	-1,102	,270	
FACTOR 1 REA.	209,000	804,000	-1,321	,187	
FACTOR 2 IMP.	278,000	873,000	-,223	,823	
FACTOR 2 REA.	216,500	811,500	-1,170	,242	
FACTOR 3 IMP.	235,500	830,500	-,767	,443	
FACTOR 3 REA.	156,500	751,500	-2,162	,031	0,0935 MEDIO
FACTOR 4 IMP.	197,500	792,500	-1,845	,065	
FACTOR 4 REA.	259,000	854,000	-,273	,785	
FACTOR 5 IMP.	278,000	431,000	-,228	,820	
FACTOR 5 REA.	261,000	397,000	-,236	,813	
FACTOR 6 IMP.	179,000	774,000	-2,262	,024	0,1023 MEDIO
FACTOR 6 REA.	208,000	803,000	-1,380	,168	

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Planificación estratégica de la evaluación; F2. Regulación del aprendizaje y de la enseñanza por medio de la evaluación; F3. Metaevaluación; F4. Ajustes en la evaluación y participación discente; F5. Adecuación y precisión; F6. Puesta en práctica de la evaluación. La valoración de la realización ligada a este último factor no se considera por la escasa consistencia interna detectada al analizar su fiabilidad.

Tabla 103

Diferencias significativas en las valoraciones de la muestra docente en función del CENTRO: estadísticos de grupo

Factores	Centro	n	M	s	SEM
F3. Metaevaluación REA.	BAM	34	2,2353	,87050	,14929
	UD	15	2,7333	,65707	,16966
F6. Puesta en práctica IMP.	BAM	34	3,8824	,61618	,10567
	UD	17	4,2353	,68733	,16670

El contraste de promedios en función del **curso** también da lugar a resultados significativos dentro del intervalo de confianza establecido. Como ya ha sido expuesto al presentar la población, el encargo docente en ambos centros se distribuye habitualmente en cursos diferentes, de modo que una misma persona puede impartir simultáneamente módulos o asignaturas en primero, segundo y/o tercero. En este contexto, el curso detalla únicamente el nivel en el que se oferta el módulo/asignatura en el que se ha realizado la evaluación.

Tabla 104

*Autovaloración de la competencia evaluativa en la muestra docente en función del CURSO***Estadísticos de contraste (H Kruskal-Wallis)**

Factores	Chi-cuadrado	gl	Sig. asintót.	TE (η^2)
Total IMPORTANCIA	3,282	2	,194	
Total REALIZACIÓN	,787	2	,675	
FACTOR 1 IMP.	,950	2	,622	
FACTOR 1 REA.	6,419	2	,040	0,1284 MEDIO
FACTOR 2 IMP.	,691	2	,708	
FACTOR 2 REA.	,372	2	,830	
FACTOR 3 IMP.	3,394	2	,183	
FACTOR 3 REA.	2,157	2	,340	
FACTOR 4 IMP.	,289	2	,865	
FACTOR 4 REA.	,705	2	,703	
FACTOR 5 IMP.	11,339	2	,003	0,2268 GRANDE
FACTOR 5 REA.	,330	2	,848	
FACTOR 6 IMP.	5,143	2	,076	
FACTOR 6 REA.	3,613	2	,164	

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Planificación estratégica de la evaluación; F2. Regulación del aprendizaje y de la enseñanza por medio de la evaluación; F3. Metaevaluación; F4. Ajustes en la evaluación y participación discente; F5. Adecuación y precisión; F6. Puesta en práctica de la evaluación. La valoración de la realización ligada a este último factor no se considera por la escasa consistencia interna detectada al analizar su fiabilidad.

Como recoge la tabla 104, la comparación de los promedios en los tres subgrupos dentro de la muestra revela una diferencia significativa ($p = .003$) y de efecto grande ($\eta^2 = 0,2268$) al valorar la relevancia de la Adecuación y la precisión en la evaluación (F5). Las opciones de contraste entre grupos que ofrece la alternativa paramétrica del análisis de varianza (ANOVA) ha permitido determinar que las diferencias tienen lugar entre el profesorado que ha contextualizado su evaluación en módulos/asignaturas de primer curso y el que lo ha hecho en segundo y tercero (ver tabla 105). En ambos casos, las puntuaciones mayores se concentran en el subgrupo docente que ha respondido desde el primer curso.

Tabla 105

Diferencias significativas en las valoraciones de la muestra docente en función del CURSO: estadísticos de grupo y contraste post hoc (Scheffé y DMS)

Diferencias significativas en función del CURSO		<i>n</i>	<i>M</i>	<i>s</i>	Sig. 1º curso	Sig. 2º curso	Sig. 3º curso
REA. FACTOR 1	1º Curso	22	3,6250	,72682	-	0,040	0,033
	2º Curso	9	3,0556	,84574	0,040	-	0,712
	3º Curso	19	3,1579	,52182	0,033	0,712	-
IMP. FACTOR 5	1º Curso	22	4,2045	,61060	-	0,026	0,009
	2º Curso	9	3,3333	,96825	0,026	-	0,958
	3º Curso	20	3,4250	,86260	0,009	0,958	-

Nota: Las celdas sombreadas señalan los grupos entre los que tiene lugar la diferencia de promedios dentro del intervalo de confianza establecido ($p < .05$).

Las diferencias a la hora de valorar el grado de cumplimiento autopercebido para el Factor 1 (Planificación estratégica de la evaluación) se producen entre los mismos grupos y en la misma dirección, aunque con una magnitud de efecto algo menor ($\eta^2 = 0,1284$) (tabla 104 y tabla 105). El profesorado que ha valorado sistemas evaluativos para módulos/asignaturas de segundo o tercer año refiere un cumplimiento adecuado en este sentido, mientras que en el curso inferior el promedio se aproxima más a la próxima categoría ($M = 3,62$, entre *adecuado* y *muy adecuado*).

En la interpretación de estos resultados es preciso considerar que, cuanto menor es el curso, mayor es la probabilidad de que la formación previa del profesorado sea de carácter pedagógico o psicopedagógico. Esta concentración se produce por efecto del Real Decreto 1393/2007, que establece que las materias y asignaturas vinculadas al módulo de formación básica (un mínimo de 60 ECTS, vinculadas principalmente a la capacitación psicopedagógica general) deben cursarse durante la primera mitad del grado. En la muestra configurada así se constata, de modo que el 72,7% del subgrupo docente del primer curso lo conforman profesorado titulado en Pedagogía, Psicología o Psicopedagogía. En consecuencia, los resultados revelados en este caso pueden guardar relación con las conclusiones extraídas al considerar el efecto potencial de la formación previa en la competencia evaluativa autopercebida. Este nexo tendría especial incidencia en la diferencia de promedios apreciada en la realización del primer factor, también señalada al analizar la titulación profesional del profesorado como criterio de agrupación.

2.2.4 Autopercepción de competencia, formación evaluativa y valoraciones docentes.

El último bloque de variables incluye la autopercepción de competencia evaluativa y dos cuestiones más asociadas a la formación permanente sobre la evaluación del aprendizaje (la realización o no de la misma y la percepción de necesidad). Como se desarrolla a continuación, en los tres casos se detecta la incidencia de diferencias estadísticamente significativas ($p < .05$) en las valoraciones de importancia y cumplimiento realizadas por el profesorado para describir su praxis evaluativa.

La **competencia evaluativa autopercebida** aparece planteada en el cuestionario docente como una variable *politómica* con tres posibles valores (poco competente, competente y muy competente). Comparadas las respuestas producidas en los tres grupos por medio de la *H* de Kruskal Wallis, no se aprecian diferencias significativas dentro del nivel de confianza asumido ($p \geq .05$). Detectado que en el grupo de alta competencia concentra un número de casos muy reducido, en una segunda fase se han contrastado los promedios de respuesta en dos subgrupos dentro de la muestra: el profesorado que se declara competente o muy competente en materia evaluativa y aquel otro que se define como poco competente en este terreno. Los resultados correspondientes a este último análisis se presentan en la tabla 106:

Tabla 106

Promedios de valoración en la muestra docente en función de la COMPETENCIA EVALUATIVA AUTOPECIBIDA (poco competente y competente/muy competente)

Estadísticos de contraste (U de Mann Whitney)

Factores	U de Mann-Whitney	W de Wilcoxon	Z	Sig. asintót. (bilateral)	TE (η^2)
Total IMPORTANCIA	115,500	935,500	-,147	,883	
Total REALIZACIÓN	71,500	92,500	-1,521	,128	
FACTOR 1 IMP.	91,500	112,500	-,941	,347	
FACTOR 1 REA.	51,500	72,500	-2,203	,028	0,0971 MEDIO
FACTOR 2 IMP.	91,000	112,000	-,960	,337	
FACTOR 2 REA.	111,500	132,500	-,185	,853	
FACTOR 3 IMP.	93,500	913,500	-,874	,382	
FACTOR 3 REA.	99,000	120,000	-,606	,544	
FACTOR 4 IMP.	95,000	915,000	-,823	,410	
FACTOR 4 REA.	105,500	126,500	-,387	,699	
FACTOR 5 IMP.	107,500	128,500	-,422	,673	
FACTOR 5 REA.	57,500	78,500	-2,057	,040	0,08476 MEDIO
FACTOR 6 IMP.	117,000	937,000	-,101	,919	

Estadísticos de contraste (*U* de Mann Whitney)

Factores	<i>U</i> de Mann-Whitney	<i>W</i> de Wilcoxon	<i>Z</i>	Sig. asintót. (bilateral)	TE (η^2)
FACTOR 6 REA.	69,500	90,500	-1,659	,097	

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Planificación estratégica de la evaluación; F2. Regulación del aprendizaje y de la enseñanza por medio de la evaluación; F3. Metaevaluación; F4. Ajustes en la evaluación y participación discente; F5. Adecuación y precisión; F6. Puesta en práctica de la evaluación. La valoración de la realización ligada a este último factor no se considera por la escasa consistencia interna detectada al analizar su fiabilidad.

La prueba de contraste *U* de Mann Whitney ha puesto de manifiesto dos diferencias de promedios con significación estadística y un efecto estimado de tamaño medio. Ambas se producen al declarar niveles de cumplimiento, la primera con respecto a la Planificación estratégica de la evaluación y la segunda en lo que refiere a su Adecuación y precisión. El profesorado que se siente menos competente con respecto a la evaluación reconoce una realización 0,7 puntos menor en los dos factores (ver tabla 107).

Tabla 107

Diferencias significativas en las valoraciones de la muestra docente en función de la COMPETENCIA EVALUATIVA AUTOPERCIBIDA

Factores	Autopercepción	<i>n</i>	<i>M</i>	<i>s</i>	<i>SEM</i>
F1. Planificación REA.	Poco competente	6	2,7083	,64064	,26154
	Competente / Muy Competente	39	3,4423	,70585	,11303
F5. Adecuación... IMP.	Poco competente	6	1,9167	,49160	,20069
	Competente / Muy Competente	39	2,6795	,89212	,14285

La realización de **formación permanente sobre evaluación del aprendizaje en los últimos dos cursos**⁹³ se ha empleado en este caso como variable dicotómica para el análisis de las valoraciones de la muestra docente en los seis factores cuestionados. Las puntuaciones de quienes han llevado a cabo alguna actividad formativa relacionada con la evaluación del aprendizaje y los que no lo han hecho difieren dentro del intervalo de confianza fijado en tres valoraciones (ver tabla 108).

⁹³ En el análisis del último objetivo profundizaremos sobre el carácter y los rasgos principales de dicha formación.

Tabla 108

Autovaloración de la competencia evaluativa en la muestra docente en función de la FORMACIÓN PERMANENTE EN EVALUACIÓN

Estadísticos de contraste (U de Mann Whitney)

Factores	U de Mann-Whitney	W de Wilcoxon	Z	Sig. asintót. (bilateral)	TE (η^2)
Total IMPORTANCIA	153,500	933,500	-1,790	,073	
Total REALIZACIÓN	120,500	900,500	-2,204	,028	0,0972 MEDIO
FACTOR 1 IMP.	149,500	929,500	-1,899	,058	
FACTOR 1 REA.	186,000	966,000	-,673	,501	
FACTOR 2 IMP.	166,500	946,500	-1,524	,128	
FACTOR 2 REA.	139,000	919,000	-1,793	,073	
FACTOR 3 IMP.	196,000	937,000	-,735	,462	
FACTOR 3 REA.	157,000	898,000	-1,261	,207	
FACTOR 4 IMP.	131,000	911,000	-2,308	,021	0,1065 MEDIO
FACTOR 4 REA.	130,500	910,500	-1,984	,047	0,0787 MEDIO
FACTOR 5 IMP.	223,000	1003,000	-,253	,800	
FACTOR 5 REA.	134,500	914,500	-1,936	,053	
FACTOR 6 IMP.	188,000	968,000	-1,051	,293	
FACTOR 6 REA.	185,500	965,500	-,704	,481	

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Planificación estratégica de la evaluación; F2. Regulación del aprendizaje y de la enseñanza por medio de la evaluación; F3. Metaevaluación; F4. Ajustes en la evaluación y participación discente; F5. Adecuación y precisión; F6. Puesta en práctica de la evaluación. La valoración de la realización ligada a este último factor no se considera por la escasa consistencia interna detectada al analizar su fiabilidad.

El subgrupo del profesorado que ha llevado a cabo formación en materia evaluativa ofrece puntuaciones significativamente mayores al valorar el grado de importancia y realización del factor cuarto, ligado a los Ajustes en la evaluación y la participación discente ($p = .021$ y $p = .047$, respectivamente). Las respuestas de quienes en los dos últimos años no han realizado actividades formativas en este espacio (el 78% de la muestra) se sitúan medio punto por debajo en ambas valoraciones (ver tabla 109). Como será expuesto al valorar los resultados relativos al objetivo H, los temas de formación permanente detallados no responden directamente a la diversificación evaluativa o a las fórmulas participativas para su desarrollo. En consecuencia, deducimos que las mayores puntuaciones obtenidas en este factor no se deben directamente a la temática abordada en la formación, sino más bien a la generación de una mayor sensibilidad o a una mayor deseabilidad social.

Tabla 109

Diferencias significativas en las valoraciones de la muestra docente en función de la FORMACIÓN PERMANENTE EN EVALUACIÓN: estadísticos de grupo

Factores	Formación	n	M	s	SEM
Total REALIZACIÓN	SI	11	3,1818	,41865	,12623
	NO	39	2,8599	,52770	,08450
F4. Ajustes/ particip. IMP.	SI	12	4,0556	,66793	,19281
	NO	39	3,5256	,72274	,11573
F4. Ajustes/ particip. REA.	SI	11	3,0000	,74162	,22361
	NO	39	2,5064	,72431	,11598

Con un tamaño de efecto medio ($\eta^2=0,0972$) y un nivel de significación dentro del intervalo de confianza fijado ($p = .028$), la diferencia de promedios en el total de la subescala de realización responde a ese mismo patrón (tabla 108). Aunque los límites de la investigación no permitan establecer relaciones de causalidad directa, la tabla 109 revela que el cumplimiento autopercebido en los indicadores de la competencia evaluativa resulta mayor en el grupo que ha participado en acciones formativas.

Adoptada como variable independiente y dicotómica, la **autopercepción de necesidad formativa con relación a la evaluación del aprendizaje** representa otro factor a explorar en las respuestas de la muestra. La prueba de contraste de la *U* de Mann Whitney revela que se produce una diferencia significativa de tamaño medio entre los dos subgrupos docentes, esto es, entre quienes no manifiestan necesidad formativa en materia evaluativa y quienes declaran sentirla (ver tabla 110). Ésta afecta a la estimación de importancia en el factor tercero, relativo a la Metaevaluación:

Tabla 110

Autovaloración de la competencia evaluativa en la muestra docente en función de la AUTOPERCEPCIÓN DE NECESIDAD FORMATIVA

Estadísticos de contraste (*U* de Mann Whitney)

Factores	<i>U</i> de Mann-Whitney	<i>W</i> de Wilcoxon	<i>Z</i>	Sig. asintót. (bilateral)	TE (η^2)
Total IMPORTANCIA	139,000	184,000	-,866	,386	
Total REALIZACIÓN	146,000	887,000	-,677	,499	
FACTOR 1 IMP.	160,000	205,000	-,301	,763	
FACTOR 1 REA.	160,500	205,500	-,287	,774	
FACTOR 2 IMP.	160,000	901,000	-,303	,762	
FACTOR 2 REA.	129,500	870,500	-1,138	,255	
FACTOR 3 IMP.	96,000	141,000	-2,050	,040	0,0840 MEDIO
FACTOR 3 REA.	152,500	197,500	-,507	,612	

Estadísticos de contraste (U de Mann Whitney)

Factores	U de Mann-Whitney	W de Wilcoxon	Z	Sig. asintót. (bilateral)	TE (η^2)
FACTOR 4 IMP.	161,000	902,000	-,273	,785	
FACTOR 4 REA.	125,000	866,000	-1,253	,210	
FACTOR 5 IMP.	128,000	173,000	-1,206	,228	
FACTOR 5 REA.	140,000	881,000	-,866	,386	
FACTOR 6 IMP.	141,500	882,500	-,822	,411	
FACTOR 6 REA.	156,500	201,500	-,409	,683	

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Planificación estratégica de la evaluación; F2. Regulación del aprendizaje y de la enseñanza por medio de la evaluación; F3. Metaevaluación; F4. Ajustes en la evaluación y participación discente; F5. Adecuación y precisión; F6. Puesta en práctica de la evaluación. La valoración de la realización ligada a este último factor no se considera por la escasa consistencia interna detectada al analizar su fiabilidad.

La ampliación de estos resultados con los estadísticos de grupo (ver tabla 111) sugiere que el profesorado que manifiesta necesidad formativa en materia evaluativa otorga mayor relevancia a la Metaevaluación. El profesorado que no cree necesitar formación califica este factor como *moderadamente relevante* ($M = 3,18$), mientras que en el otro grupo se produce una puntuación media muy próxima al 4 (*relevante*).

Tabla 111

Diferencias significativas en las valoraciones de la muestra docente en función de la NECESIDAD FORMATIVA AUTOPECIBIDA: estadísticos de grupo

Factores	Necesidad	n	M	s	SEM
F3. Metaevaluación IMP.	SÍ	38	3,9561	,88759	,14399
	NO	9	3,1852	,97341	,32447

El último de los objetivos vinculado al marco empírico (H) profundiza en la descripción de estas dos últimas variables desde una visión cualitativa.

Breve avance de la discusión de resultados...

La complejidad y singularidad de los procesos formativos hace que la docencia y la evaluación estén abiertas a una amplia variedad de influencias potenciales. De las diferencias y correlaciones puestas de manifiesto, resultan especialmente destacables las relativas a la formación previa y permanente del profesorado, sobre las que el próximo capítulo prestará particular atención.

En contra de lo evidenciado en otros estudios similares (Caballero, 2013; Palacios & López, 2013) la

formación previa parece incidir más y con mayor efecto que la formación permanente en el cumplimiento declarado para los factores asociados a la competencia evaluativa docente. Reabriendo el clásico debate sobre los modelos de formación docente simultáneos y consecutivos y la falta de capacitación psicopedagógica específica en el acceso a la docencia universitaria (J. M. Esteve, 2006; I. Fernández & Márquez, 2014; Montero, 2006), en la muestra analizada los mayores promedios de realización se concentran en el colectivo de tituladas/os en Pedagogía, Psicología y Psicopedagogía. Esta conclusión no encuentra respaldo unánime en los resultados publicados en otros trabajos; junto a testimonios que apoyan la tesis de la influencia de la formación previa en las prácticas evaluativas (V. Álvarez et al., 2002; Gil-Flores, 2012; Jessop & Maleckar, 2014; N. Martínez et al., 2012; Tejedor, 1999), se constatan otros que la desmienten (Porto, 2006; Serrano, 2010).

Las valoraciones puestas de manifiesto en la muestra discente (objetivo F) ofrecen ciertos indicios que podrían reafirmar la incidencia potencial de la formación previa docente en la praxis evaluativa. Así se constata al observar mejores resultados de realización donde mayor es la concentración de tituladas/os en Pedagogía, Psicología o Psicopedagogía (módulo de formación básica y primer curso). Las demandas formativas del profesorado (objetivo H) también sugieren la necesidad de construir una base común para el ejercicio de la evaluación, clarificando desde la perspectiva psicopedagógica (fases del proceso, tipología y definiciones básicas, etc.).

Aunque con menor efecto que en la formación previa, también han sido detectadas diferencias significativas en función de la realización de formación permanente en evaluación. Este resultado reafirma una de las principales líneas de mejora apuntadas desde la literatura y la investigación específica, que sitúan la formación evaluativa del profesorado en ejercicio como medio ineludible para promover el desarrollo de la competencia evaluativa (V. Álvarez et al., 2002; Ibarra & Rodríguez, 2010; Nieto, 2000; Palacios & López, 2013).

2.3 Resultados Objetivo E3. Identificar las principales fortalezas, debilidades y dificultades para el ejercicio de la evaluación desde el punto de vista docente.

El tercer sub-propósito completa las conclusiones obtenidas en el estudio inferencial y descriptivo de la competencia evaluativa relatado en los dos primeros. Parte para ello de las respuestas ofrecidas por el profesorado de la muestra a una pregunta final abierta sobre los principales obstáculos y apoyos que acompañan al ejercicio de la praxis evaluativa. Respondida por el 84,4% de las/os participantes, el tono general de las aportaciones recibidas en la misma resulta predominantemente negativo según detalla la figura 40:

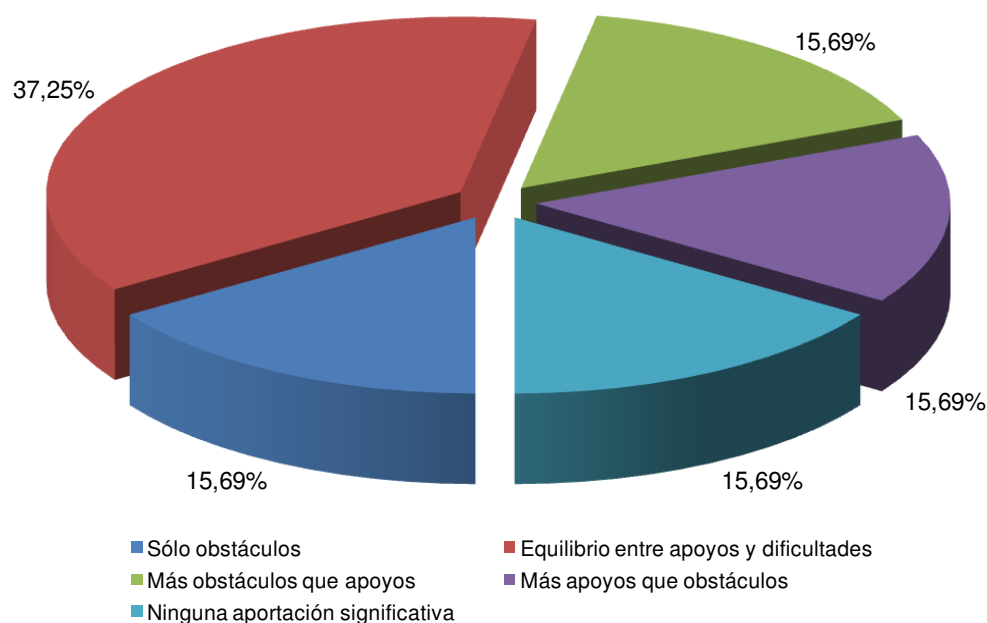


Figura 40. Resumen del tono general de las respuestas docentes en la pregunta referida a los apoyos y obstáculos en el ejercicio de la competencia evaluativa.

Las/os 43 docentes que han respondido a esta cuestión han realizado un total de 145 aportaciones, 54 de ellas referidas a apoyos para el cumplimiento de sus funciones evaluativas (37,24%) y las 91 restantes (62,75%) identificando obstáculos y dificultades. El número de respuestas o aportaciones no es limitado ni excluyente, de modo que los datos que se ofrecen a continuación parten de dicho total.

Las respuestas a la primera parte de la pregunta (los apoyos) han sido categorizadas en nueve núcleos recogidos en la tabla 112. Entre ellos, los recursos de apoyo más mencionados son la *planificación rigurosa* y el *apoyo e intercambio de experiencias con colegas* ($f = 10$; 18,5% de las aportaciones, respectivamente). En la primera categoría, las descripciones recibidas refieren al diseño sistemático del sistema evaluativo, incluyendo el claro establecimiento de los indicadores que se adoptan como criterio y como meta y la creación de recursos específicos para su control y difusión (rúbricas, listas de control, etc.). Son frecuentes en este sentido las alusiones al proceso completo de planificación de la guía de aprendizaje de la asignatura, donde la evaluación representa un capítulo destacado. En el *apoyo de colegas e intercambio de experiencias*, el profesorado consultado subraya las consultas, los contrastes o los debates y discusiones con otros miembros del equipo docente como un recurso que facilita la evaluación del aprendizaje. También se señala la *colaboración de expertos y las actividades formativas* como otra forma de apoyo externo más técnica y específica. Casi el 13% de las ayudas sugeridas recaen dentro de dicha categoría, proporción que se repite para el apoyo

denominado *revisión anual e introducción de ajustes*. En este caso, las respuestas describen la mejora y revisión continua de los sistemas evaluativos aplicados de un año a otro, de modo que se van introduciendo cambios y ajustes en función de la experiencia y los resultados constatados.

Tabla 112

Apoyos para el ejercicio de la competencia evaluativa según las respuestas abiertas del profesorado

APOYOS y AYUDAS: categorías de respuesta	f	%
Planificación rigurosa (indicadores claros, claridad, instrumentos de control, etc.)	10	18,50
Apoyo de colegas e intercambio experiencias	10	18,50
Apoyo experto y actividades formativas	7	12,96
Revisión anual e introducción de ajustes (ensayo error, investigación-acción, etc.)	7	12,96
Consulta del punto de vista de estudiantes sobre la evaluación (encuesta, feedback, comentarios y sugerencias de mejora, etc.)	6	11,11
Evaluación continua de las/os estudiantes (análisis de resultados parciales)	6	11,11
Experiencia y dominio de la materia	4	7,40
Trabajo en grupo y actividades de evaluación entre compañeras/os para reducir la carga de la corrección y el feedback	3	5,55
Plataforma ALUD	1	1,85
TOTAL DE LOS APOYOS IDENTIFICADOS	54	

Las respuestas recopiladas sugieren que el alumnado representa otro apoyo para el ejercicio de las funciones evaluativas. Su efecto se explica por dos vías complementarias, que suman algo más del 22% de las ayudas reseñadas. Por una parte, el profesorado señala la contribución que suponen los *comentarios y observaciones que realiza el alumnado* sobre el sistema de evaluación, ya sea directamente o por medio de las encuestas de evaluación de la calidad docente. Por otra parte, la propia *evaluación continua de las/os estudiantes* se indica como una fuente de información que conduce a la mejora del planteamiento evaluativo, al revelar sus fortalezas y deficiencias antes de la evaluación final. La adecuación del nivel de exigencia, la pertinencia de las tareas planteadas o los riesgos y limitaciones de las herramientas escogidas representan algunas de las cuestiones que pueden valorarse a partir de los resultados de esa evaluación continua, según las respuestas registradas.

Con menor representación en el banco de respuestas recopilado, el profesorado de la muestra señala otras ayudas como la *propia experiencia y el dominio de la materia* (7,40%) y la adopción de medidas para reducir la carga que genera la evaluación (*fórmulas*

participativas de evaluación que reducen la carga de corrección y devolución o el uso de la *plataforma ALUD*, herramienta de trabajo basada en Moodle).

En cuanto a los obstáculos y dificultades que median en la evaluación del aprendizaje, la categorización propuesta es más compleja y discrimina 10 temas principales (ver tabla 113).

Tabla 113
Obstáculos para el ejercicio de la competencia evaluativa según las respuestas abiertas del profesorado

OBSTÁCULOS y DIFICULTADES	f	%
SOBRECARGA	40	43,95
⇒ Número elevado de estudiantes	23	25,27
⇒ Sobrecarga profesional (volumen de trabajo, otras funciones, etc.)	9	9,88
⇒ Condiciones prácticas (espacios, plazos, convocatorias, soporte técnico)	4	4,39
⇒ Sobrecarga generada por el sistema de evaluación	4	4,39
ATENCIÓN INDIVIDUALIZADA	19	20,87
⇒ Seguimiento individualizado y comisión de feedback	12	13,18
⇒ Dificultad para aplicar métodos de evaluación más personalizados	7	7,69
FALTA DE FORMACIÓN EN EVALUACIÓN	17	18,68
⇒ Inseguridad, falta de formación en procedimientos de evaluación (técnicas, principios, herramientas, desconocimiento de buenas prácticas)	11	12,08
⇒ Complejidad de la evaluación de competencias	6	6,59
DIFICULTADES ASOCIADAS AL GRUPO DE ESTUDIANTES	4	4,39
⇒ Presencia irregular del alumnado en el aula	3	3,29
⇒ Falta de hábito del alumnado en sistemas de evaluación de competencias	1	1,09
Exigencias curriculares	2	2,19
Evaluación del trabajo realizado en grupos	2	2,19
Falta de coordinación entre el profesorado en la evaluación	2	2,19
Rigidez por la publicación de compromisos evaluativos en la guía de aprendizaje	2	2,19
Inexperiencia docente	2	2,19
Otros factores personales	1	1,09
TOTAL DE OBSTÁCULOS IDENTIFICADOS	171	

Nota: Las 10 categorías principales están reseñadas en negrita.

La más evidente de las dificultades manifestada por el profesorado de la muestra es la *sobrecarga*, que concentra el 43,95% de las 91 aportaciones registradas. Esta primera categoría integra obstáculos de naturaleza diversa que dificultan el desempeño de las funciones evaluativas según las/os propias/os docentes. El *elevado número de estudiantes* representa la condición más reiterada en este sentido ($f = 23$), aunque no es la única. La

cantidad y variedad de funciones que debe desempeñar el profesorado universitario con respecto a la docencia, la gestión, la investigación o la transferencia (*sobrecarga profesional*) también figura como obstáculo en las descripciones recopiladas. A ella se suma la *sobrecarga ligada a los propios sistemas de evaluación articulados* (acumulación de correcciones durante el proceso, dificultades para devolver feedback a tiempo, etc.) y las condiciones prácticas en las que éstos se ponen en juego que, a menudo, agregan dificultades (gestión de las convocatorias, plazos de publicación de resultados, rigidez del calendario académico, etc.).

La imposibilidad de ofrecer al alumnado una *atención individualizada en la evaluación* figura como la segunda categoría más señalada entre las dificultades (20,87%). En las respuestas recopiladas este factor aparece estrechamente relacionado con el número de estudiantes y adopta dos manifestaciones complementarias: el 13,18% de las aportaciones señalan grandes dificultades para hacer *seguimiento individualizado* del alumnado y ofrecerle feedback personal sobre sus logros y su evolución y el 7,69% refiere a la imposibilidad de adoptar *métodos de evaluación más flexibles e individualizados* por problemas de viabilidad práctica.

En las aportaciones ofrecidas por el profesorado *la falta de formación específica en evaluación* se presenta como otra dificultad añadida para su ejercicio. Los testimonios recabados en este sentido describen *inseguridad y/o falta de conocimiento* con respecto a los principios, métodos o instrumentos aplicables ($f = 11$) a la praxis evaluativa; también apuntan a la *mayor complejidad que supone la evaluación de competencias* frente a la clásica evaluación basada en contenidos ($f = 6$).

El resto de factores recogidos en la tabla 113 tienen un reflejo mucho menor entre los obstáculos detallados en la muestra. Las dificultades que se asocian al propio grupo de estudiantes no llegan al 5% de las aportaciones e incluyen su *falta de hábito con respecto a la evaluación de competencias*, así como las dificultades que ocasiona la *asistencia irregular en sistemas que exigen presencialidad*. Para una pequeña parte del profesorado consultado, la publicación de las guías de aprendizaje y los compromisos evaluativos representa una dificultad al ocasionar *rigidez* (2,19% de las respuestas). En la misma medida aparecen representados otros obstáculos como la propia *inexperiencia docente*, las *exigencias curriculares* derivadas de los planes de estudio (desacuerdo con las competencias adjudicadas y los resultados de aprendizaje asociados), la *descoordinación del profesorado en materia evaluativa* o las clásicas *dificultades para valorar el trabajo realizado en grupos* (aportaciones desiguales entre las/os integrantes, conflictos, etc.).

Breve avance de la discusión de resultados...

La valoración global que realiza el profesorado de los apoyos y las dificultades en el ejercicio de la evaluación reafirma las conclusiones comunicadas en otras investigaciones similares. Entre las aportaciones recopiladas se evidencian dos condiciones que adoptan una especial significancia en el ejercicio de las funciones evaluativas. La primera de ellas hace referencia a las condiciones en las que transcurre el trabajo docente, mientras que la segunda apela a la formación permanente del profesorado en materia evaluativa.

Como en otros estudios centrados en la praxis evaluativa (L. F. Martínez et al., 2012; Porto, 2006; Tejedor, 1999) la sobrecarga (particularmente la asociada al elevado número de estudiantes) representa la mayor de las dificultades reconocidas. En opinión del profesorado, su incidencia provoca un insuficiente nivel de atención y seguimiento individualizado en la evaluación (Zaragoza et al., 2009), ocasionando un efecto negativo sobre los procesos de retroalimentación. Este perjuicio resulta especialmente alarmante ya que el feedback representa el más potente de los recursos formativos que proporciona la evaluación. Así se constata ampliamente desde la perspectiva teórica (Black & Wiliam, 1998b; Evans, 2013; Hattie & Timperley, 2007; Nicol et al., 2014), y así lo aprecian también el alumnado y profesorado de la muestra (objetivos E1 y F1).

En cuanto a la formación permanente en materia evaluativa, las respuestas recopiladas sugieren en ella una alternativa de mejora prioritaria directa e indirectamente. Por una parte, las respuestas recopiladas refieren a la dificultad que ocasiona la insuficiente formación o capacitación inicial en esta competencia, confirmando las apreciaciones de Stiggins (1993/2012, 1998) o B. Álvarez et al. (2007). Por otra parte, entre los apoyos para su desempeño se constatan referencias a alternativas formativas basadas en el intercambio horizontal de experiencias y en la discusión de casos y problemas reales, en línea con las recomendaciones de Caballero (2013), Blanch y Corcelles (2014), Ibarra y Rodríguez (2010), Mateo (2012) o Torra et al. (2013). Frente a los clásicos cursos, esta modalidad tiene mayor probabilidad de provocar cambios en las concepciones evaluativas profundas y es señalada como predilecta por el profesorado (objetivo H).

3. Resultados relativos al Objetivo F. La competencia evaluativa docente desde la perspectiva del alumnado

El segundo objetivo del marco empírico analiza la valoración que realiza el alumnado de la competencia evaluativa del profesorado, según el nivel de importancia y cumplimiento atribuido a los indicadores cuestionados. Su abordaje se ha efectuado por medio de tres

sub-propósitos complementarios, abordados en los próximos apartados: describir las respuestas producidas en la muestra discente, indagar sobre la posible generalización de las diferencias de promedios advertidas en función de las variables de identificación consideradas y analizar las principales fortalezas y debilidades que se aprecian en la práctica evaluativa desde la perspectiva del alumnado.

3.1 Resultados Objetivo F1. Identificar los factores e indicadores de la evaluación del aprendizaje en los que mayor y menor dificultad acusa el alumnado, atendiendo al nivel de relevancia y realización que les atribuyen.

El análisis factorial de la respuesta discente ha identificado seis factores implicados en la competencia evaluativa y a ellos refieren las valoraciones de relevancia y realización que se presentan en este nuevo apartado. Expuestos en el capítulo anterior, la definición de dichos factores no coincide con la estructuración interna detectada en las respuestas del profesorado, ni con la delimitación inicial de dimensiones planteada en el modelo teórico de referencia. Así lo constata el **ANEXO 27**, que ofrece un contraste de las tres estructuras resultantes con sus coincidencias y discrepancias.

El alumnado de la muestra considera *relevantes* los componentes asociados a la competencia evaluativa, sin grandes discrepancias de opinión ($M = 4,29$; $s = 0,37$). A la hora de valorar su grado de realización, en cambio, aumenta la dispersión de las respuestas y el promedio de valoración se sitúa por debajo de la media teórica ($M = 2,82$; $s = 0,61$). Estas valoraciones globales muestran un alto grado de concordancia con las efectuadas por la muestra docente que, al igual que el alumnado, sitúa la competencia evaluativa en la zona de Concentración de la rejilla AIR.

A juzgar por los promedios de respuesta registrados (ver tabla 114), las/os estudiantes de la muestra considera *relevantes* todos y cada uno de los factores considerados en la valoración de la competencia evaluativa ($M \geq 4$ en todos los casos). Igual que en la muestra docente, la Regulación del aprendizaje por medio de la evaluación (F4) destaca como el componente más importante para el alumnado; en él se concentran el mayor promedio de relevancia ($M = 4,52$) y el menor grado de dispersión ($s = 0,53$), lo que indica concordancia en las respuestas obtenidas. En el extremo opuesto, la Participación de las/os estudiantes en la evaluación (F1) figura como el menos importante de los factores considerados ($M = 4,08$, lo que lo sigue situando como un factor *relevante*).

Tabla 114

Valoración de la muestra discente en los seis factores asociados a la competencia evaluativa docente

Estadísticos descriptivos					
Factores (F)	N	Mínimo	Máximo	Media	s
F1. Participación discente. IMP.	704	1,25	5,00	4,0888	0,66212
F1. Participación discente. REA.	704	1,00	5,00	2,2232	0,95528
F2. Planificación evaluación. IMP.	704	2,20	5,00	4,2892	0,44498
F2. Planificación evaluación. REA.	704	1,40	5,00	3,0386	0,73746
F3. Ajustes evaluación. IMP.	704	1,00	5,00	4,4100	0,57560
F3. Ajustes evaluación. REA.	700	1,00	5,00	3,7988	1,02405
F4. Regulación aprendizaje. IMP.	703	1,33	5,00	4,5292	0,53203
F4. Regulación aprendizaje. REA.	641	1,00	5,00	2,3250	0,94869
F5. Adecuación y precisión. IMP.	703	1,00	5,00	4,4047	0,68678
F5. Adecuación y precisión. REA.	701	1,00	5,00	3,0342	1,02028
F6. Control condiciones. IMP.	704	1,00	5,00	4,2363	0,56968
F6. Control condiciones. REA.	703	1,00	5,00	3,4291	0,87818
N válido (según lista)	704				

Nota: La fuente azul destaca los factores con mayor promedio de puntuación en las subescalas de importancia (IMP.) y realización (REA.). La fuente roja señala los menores promedios registrados en ambos casos.

Como en la muestra docente, en la subescala de realización se produce un evidente descenso de los promedios de respuesta y el aumento de los índices de dispersión. En opinión del alumnado consultado, el factor tercero (Ajustes en el sistema evaluativo) es el más representado en la práctica; el grado de cumplimiento estimado para el mismo es entre *adecuado* y *muy adecuado* ($M = 3,79$), aunque con elevado nivel de desviación entre las respuestas ($s = 1,024$, la dispersión más acusada de toda la subescala). El Control de las condiciones evaluativas (F6), su Adecuación y precisión (F5) y la Planificación de la evaluación (F2) también se llevan a cabo de forma aceptable a juicio de las/os estudiantes ($M \geq 3$, realización *adecuada*). No sucede igual en los dos factores restantes, donde las valoraciones registradas dan cuenta de un grado de cumplimiento *poco adecuado*. La Participación de las/os estudiantes en la evaluación (F1), además de ser el factor menos valorado en relevancia, es el menos llevado a cabo según el alumnado de la muestra ($M = 2,22$; $s = 0,95$). El colectivo discente no percibe que se requiera su participación en el diseño, desarrollo o revisión de los sistemas evaluativos experimentados. Tampoco siente que la evaluación se emplee para regular su aprendizaje por medio del feedback y las sugerencias de mejora. Así lo evidencia el promedio de realización registrado para el factor cuarto (Regulación del aprendizaje por medio de la evaluación), que se sitúa claramente por

debajo de la media teórica ($M = 2,32$). Este último dato contrasta con las valoraciones de importancia registradas, que sitúan dicho factor como el más relevante.

Representados en la rejilla AIR los promedios de importancia y realización percibida registrados (ver figura 41), el único factor situado en la zona de Mantenimiento es el relativo a los Ajustes en el sistema evaluativo (F3). Las/os estudiantes entienden que la comunicación y gestión de los cambios que afectan a la evaluación representa una cuestión entre *importante* y *muy importante* ($M = 4,40$) y que el profesorado la lleva a cabo de forma más que *adecuada* ($M = 3,79$ próximo a 4, *muy adecuada*).

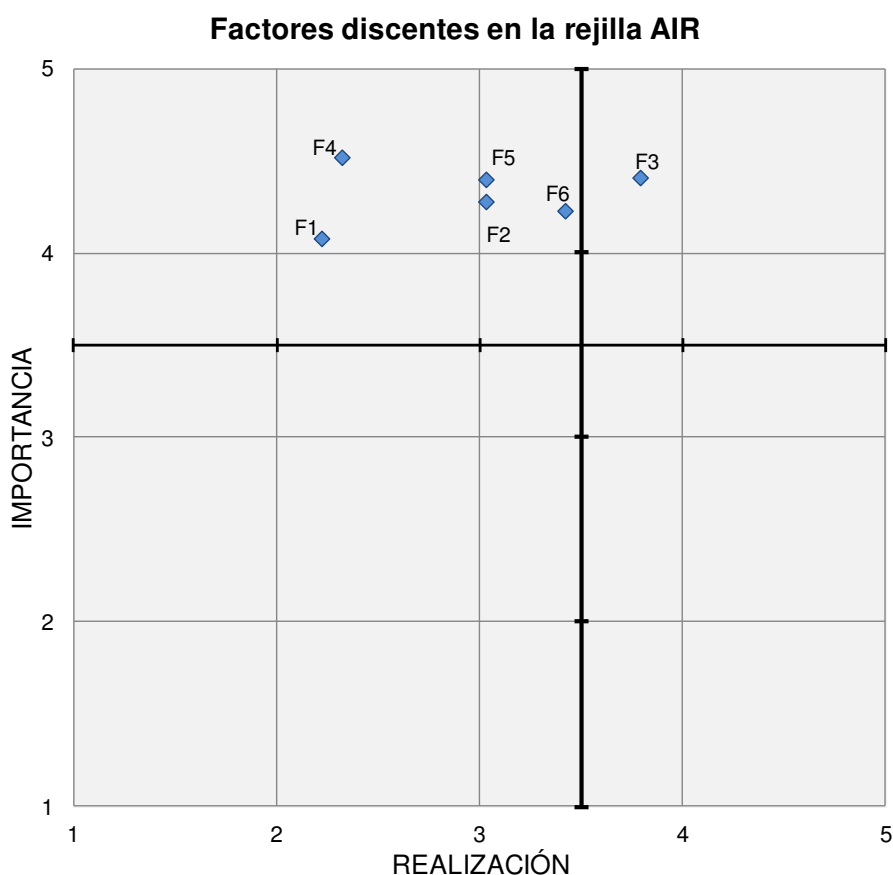


Figura 41. Representación AIR de los factores asociados a la competencia evaluativa docente en la muestra de estudiantes.

Los factores se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación.

Los cinco factores restantes se ubican en el cuadrante de Concentración, caracterizado por el desajuste entre las valoraciones de importancia ($M \leq 4$) y las de realización ($M \leq 3,5$). El factor cuarto, sobre la Regulación del aprendizaje por medio de la evaluación, se revela como el más desajustado, por presentar el mayor desfase entre los

promedios de relevancia y realización percibida (más de dos puntos de diferencia). Aunque se encuentre también en la zona de Concentración, el desfase referido al Control de las condiciones en las que se desarrolla la evaluación (F6) parece menos grave para la muestra; con un promedio de realización percibida de 3,42, este aspecto se sitúa en el límite con el área de Mantenimiento ($M \geq 3,5$).

Expuestos los principales resultados sobre los seis factores presentes en el cuestionario discente, la figura 42 resume las valoraciones de la muestra a los 20 ítems que lo integran (dicha información se presenta ampliada para su consulta en el **ANEXO 28**). El traslado de estas valoraciones a la rúbrica de valoración empleada para su recopilación, devuelve una imagen más descriptiva de la visión del alumnado sobre la competencia evaluativa de sus profesoras/es. Su consulta puede realizarse en el **ANEXO 29** como complemento de esta primera interpretación cuantitativa.

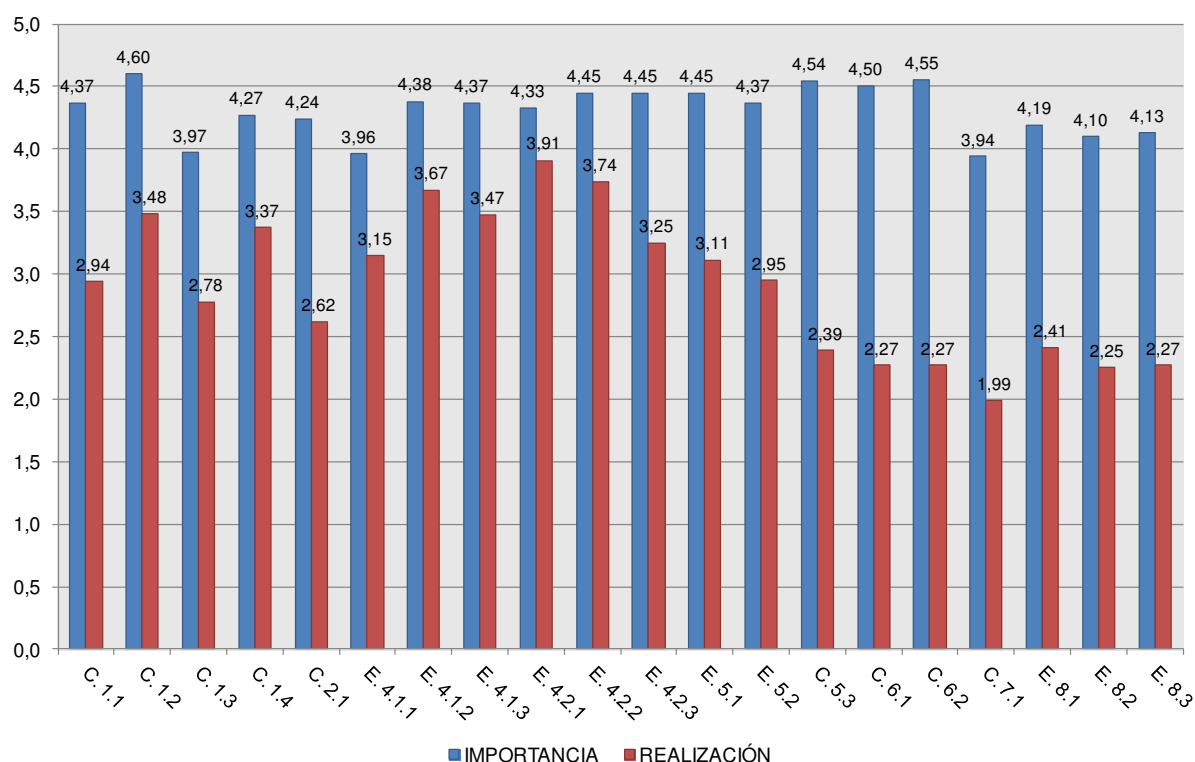


Figura 42. Promedios de importancia y realización declarados por la muestra discente en los 20 indicadores de competencia evaluativa.

De acuerdo a lo expuesto en el ANEXO 17, los ítems encabezados con C son comunes para docentes y alumnado y los precedidos por E, exclusivos para estudiantes.

En coherencia con lo expuesto hasta el momento, los datos más llamativos se concentran en torno a los factores primero y cuarto. El ítem C.7.1., el más directamente referido a *la implicación discente en la evaluación*, es el menos valorado por la muestra,

tanto en importancia, como en realización. Los tres ítems relativos a la participación del alumnado en la *metaevaluación* (E.8.1., E.8.2. y E.8.3.) completan la lista de los menos valorados en importancia y también figuran entre los que en menor medida se consideran representados en la práctica. Todos ellos conforman el primer factor (Participación de estudiantes en la evaluación), aquel con menores puntuaciones en ambas subescalas.

El indicador al que mayor importancia concede la muestra discente es el relativo a la *alineación de la evaluación con los objetivos formativos y la metodología* ($M = 4,60$). Le siguen en orden de relevancia los tres ítems que conforman el Factor 4 (Regulación del aprendizaje), todos ellos referidos a la *comunicación de resultados de aprendizaje* y al establecimiento de *procesos de feedback* y mejora o *proacción* a partir de los mismos (C.5.3., E.6.1. y E.6.2.). Sin embargo, el grado de realización que perciben las/os estudiantes para estos tres elementos es *poco adecuado* ($M \leq 2,39$); junto a los componentes del factor primero, se encuentran en el grupo de los indicadores peor valorados en la subescala.

Las valoraciones de relevancia realizadas por el alumnado consultado presentan una desviación media de 0,739, sensiblemente inferior a la constatada para el profesorado ($s = 0,849$) en una muestra mucho menor; no así en la subescala de realización, donde la dispersión media es superior al punto ($s = 1,14$) y mayor que la calculada para el sector docente ($s = 0,953$). Estos datos reflejan diferencias evidentes a la hora de valorar los factores y componentes asociados a la evaluación del aprendizaje, en especial, en lo que tiene que ver con su puesta en práctica. El análisis del siguiente objetivo, profundiza en la naturaleza y significación de tales diferencias, así como en los factores socio-demográficos, profesionales o académico institucionales que parecen mediar en las mismas.

Breve avance de la discusión de resultados...

Investigaciones próximas sitúan la evaluación del aprendizaje como una de las principales áreas de mejora en la formación universitaria (Buckley, 2012; I. González, 2003; Higher Education Founcil Council for England- HEFCE, 2014). Los datos recabados de la muestra ofrecen un testimonio similar para la formación básica de maestras/os, con discrepancias evidentes entre el modelo teórico establecido (ITC) y las prácticas evaluativas vivenciadas por el alumnado. Esta imagen general se manifiesta de forma similar en Hamodi y López (2012) y V. M. López y Palacios (2012), que desarrollan sus estudios en el mismo contexto formativo.

Los desajustes detectados resultan especialmente evidentes en dos factores: el relativo a la regulación del aprendizaje por medio de la evaluación y el referido a la intervención discente en la misma. En contra de las orientaciones desprendidas de la literatura específica y de las exigencias de

la valoración de competencias en el marco del EEES, la evaluación participativa representa para las/os estudiantes el menos importante de los factores asociado a la competencia evaluativa. Confirmando los hallazgos de otros muchos estudios, destaca a la vez como el menos llevado a cabo en la práctica (Arribas et al., 2010; Gil-Flores, 2012; G. Rodríguez et al., 2012; Ruiz-Gallardo et al., 2013; Smitter, 2006). En cuanto a la Regulación del aprendizaje por medio de la evaluación, concentra en sí un importante nivel de desajuste. A pesar de destacar como el más importante de los factores analizados, su incidencia práctica se considera insuficiente. En línea con las conclusiones reportadas en trabajos anteriores (Gutiérrez et al., 2011; Gutiérrez et al., 2013; Jessop, 2012; L. F. Martínez et al., 2012), el alumnado describe procesos de feedback infrecuentes, impuntuales e insuficientemente personalizados. El profesorado, por su parte, relaciona esta carencia con la sobrecarga (número de estudiantes, excesiva carga docente, etc.) (objetivo E3).

La posición que adopta el alumnado ante la diversificación de los procesos evaluativos deja planteada otro interrogante interesante para su discusión. Suprimido como factor por las amenazas psicométricas constatadas en el mismo, evidencias indirectas sugieren que la justicia y adecuación evaluativa se interpretan en términos de homogeneidad entre el alumnado, confirmando la tendencia apreciada en el análisis de la normativa evaluativa (Real Decreto 1791/2010).

3.2 Resultados Objetivo F2. Identificar relaciones entre distintas variables académicas, profesionales o socio-demográficas y la competencia evaluativa percibida por el alumnado.

Para indagar sobre este segundo sub-propósito se ha recurrido a pruebas de inferencia estadística bivariable. En el análisis se han explorado las valoraciones de importancia y realización arrojadas por la muestra discente, en función de 15 variables de naturaleza diferente (socio-demográfica, académica y profesional). Probablemente asociado al tamaño de la muestra ($n = 704$), los resultados obtenidos revelan un número muy elevado de diferencias significativas dentro del intervalo de confianza establecido (IC = 95%) (ver tabla 115). Asociadas a 14 de las 15 variables consideradas, la magnitud media del efecto estimado para dichas diferencias es pequeña ($\eta^2 = 0,0483^{94}$).

La única variable en la que no se han detectado diferencias estadísticamente significativas es la **Modalidad de acceso a la formación básica de maestras/os**. Según los resultados registrados, el acceso por medio de las distintas vías (bachillerato, FP,

⁹⁴ Promedio de las puntuaciones eta cuadrada de todas las diferencias estadísticamente significativas detectadas en la muestra de estudiantes.

acceso a mayores de 25 años, otra titulación universitaria, o estudios universitarios inconclusos) no parece asociado a valoraciones de relevancia o cumplimiento significativamente diferentes en los seis factores cuestionados.

Tabla 115

Conclusiones estadísticamente significativas ($p < .05$) en la valoración de la competencia evaluativa docente en la muestra discente por factor y subescala (Sub.)

FACTOR	Sub.	CONCLUSIÓN
TOTAL	IMP.	• SEXO. Las mujeres mayor promedio (TE MUY PEQUEÑO). • EDAD. Correlación positiva y muy baja. • GRUPOS EDAD. El grupo de menor edad menor promedio que el de 20 años (TE PEQUEÑO). • DEDICACIÓN FORMACIÓN. El alumnado que trabaja sin contrato mayor promedio que los que se dedican en exclusiva a la formación (TE PEQUEÑO). • GRADO. El alumnado de Educación Infantil menores promedios (TE PEQUEÑO). • CURSO. El alumnado de tercero mayor promedio (TE PEQUEÑO). • UNIDAD VALORACIÓN. En módulos mayor promedio (TE MUY PEQUEÑO). • MATERIA. Diferencias significativas (TE MEDIO).
	REA.	• SEXO. Los hombres mayor promedio (TE PEQUEÑO). • EDAD. Correlación negativa y muy baja. • GRADO. El alumnado de Educación Infantil menores promedios (TE PEQUEÑO). • CURSO. El alumnado de 1º mayor promedio que el de 2º y 3º (TE PEQUEÑO). • UNIDAD VALORACIÓN. En módulos menor promedio (TE MEDIO). • MATERIA. Diferencias significativas (TE GRANDE). • MÓDULO. El alumnado del módulo didáctico disciplinar menor promedio que el de mención y el de formación básica (TE PEQUEÑO). • RESULTADOS PARCIALES. Quienes disponen de resultados, mayor promedio. (TE PEQUEÑO) • EXPECTATIVAS. Quienes esperan aprobar mayor promedio que los que manifiestan dudas (TE PEQUEÑO). • COMPETENCIA DOCENTE PERCIBIDA. Correlación positiva y moderada. • COMPETENCIA DOCENTE PERCIBIDA. Diferencias significativas entre los tres grupos, donde los mayores promedios se concentran en el grupo de mayor competencia (TE GRANDE). • MODALIDAD. Mayor promedio en la modalidad abreviada (TE PEQUEÑO).
FACTOR 1	IMP.	• SEXO. Las mujeres mayor promedio (TE PEQUEÑO). • DEDICACIÓN FORMACIÓN. El alumnado que trabaja sin contrato mayor promedio que los que se dedican en exclusiva a la formación (TE PEQUEÑO). • CURSO. El alumnado de tercero mayor promedio que el segundo (TE PEQUEÑO). • CENTRO. El alumnado de UD mayor promedio (TE MUY PEQUEÑO). • MATERIA. Diferencias significativas (TE PEQUEÑO).

FACTOR	Sub.	CONCLUSIÓN
FACTOR 1	REA.	• SEXO. Los hombres mayor promedio (TE MUY PEQUEÑO). • EDAD. Correlación negativa y muy baja. • GRUPOS EDAD. El grupo menor, mayor promedio que los dos grupos mayores (TE PEQUEÑO). • GRADO. El alumnado de Educación Infantil menores promedios (TE PEQUEÑO). • CURSO. El alumnado de 1º mayor promedio que el de 2º y 3º (TE PEQUEÑO). • UNIDAD VALORACIÓN. En módulos menor promedio (TE MUY PEQUEÑO). • MATERIA. Diferencias significativas (TE GRANDE). • MÓDULO. El alumnado de DECA menor promedio que el resto de módulos; el del módulo didáctico disciplinar menor promedio que el de formación básica (TE PEQUEÑO). • EXPERIENCIA PREVIA. Quienes tienen experiencia previa con el profesorado de referencia mayor promedio (TE MUY PEQUEÑO). • EXPECTATIVAS. Quienes esperan aprobar mayor promedio que los que manifiestan dudas (TE PEQUEÑO). • COMPETENCIA DOCENTE PERCIBIDA. Correlación positiva y pequeña. • COMPETENCIA DOCENTE PERCIBIDA. Diferencias significativas entre los tres grupos, donde los mayores promedios se concentran en el grupo de mayor competencia (TE GRANDE). • MODALIDAD. Mayor promedio en la modalidad abreviada (TE PEQUEÑO).
	IMP.	• SEXO. Las mujeres mayor promedio. (TE PEQUEÑO) • EDAD. Correlación positiva y muy baja. • GRUPOS EDAD. El grupo menor, menor promedio que los dos grupos mayores (TE PEQUEÑO). • GRADO. El alumnado de Educación Infantil mayores promedios (TE PEQUEÑO). • CURSO. El alumnado de 3º mayor promedio que el de 1º (TE PEQUEÑO). • MATERIA. Diferencias significativas (TE PEQUEÑO). • COMPETENCIA DOCENTE PERCIBIDA. Correlación positiva y alta.
FACTOR 2	REA.	• SEXO. Los hombres mayor promedio (TE PEQUEÑO). • EDAD. Correlación negativa y muy baja. • GRUPOS EDAD. El grupo menor, mayor promedio que el alumnado de 20 años (TE PEQUEÑO). • GRADO. El alumnado de Educación Infantil menores promedios (TE PEQUEÑO). • CURSO. El alumnado de 3º menor promedio que el de 1º y 2º (TE PEQUEÑO). • CENTRO. El alumnado de BAM mayor promedio (TE MUY PEQUEÑO). • UNIDAD VALORACIÓN. En módulos menor promedio (TE MUY PEQUEÑO). • MATERIA. Diferencias significativas (TE GRANDE). • MÓDULO. El alumnado del módulo didáctico disciplinar menor promedio que el de mención y el de formación básica (TE PEQUEÑO). • RESULTADOS PARCIALES. Quienes disponen de resultado, mayor promedio (TE PEQUEÑO). • EXPECTATIVAS. Diferencias significativas (no se puede precisar grupos) (TE PEQUEÑO). • COMPETENCIA DOCENTE PERCIBIDA. Correlación positiva y moderada. • COMPETENCIA DOCENTE PERCIBIDA. Diferencias significativas entre los tres grupos, donde los mayores promedios se concentran en el grupo de mayor competencia (TE GRANDE).

FACTOR	Sub.	CONCLUSIÓN
FACTOR 3	IMP.	• SEXO. Las mujeres mayor promedio (TE MUY PEQUEÑO). • DEDICACIÓN FORMACIÓN. El alumnado que trabaja sin contrato mayor promedio que los que se dedican en exclusiva a la formación (TE MUY PEQUEÑO). • GRADO. El alumnado de Educación Infantil mayores promedios (TE MUY PEQUEÑO). • MATERIA. Diferencias significativas (TE PEQUEÑO).
	REA.	• CURSO. El alumnado de 1º menor promedio que el de 3º (TE PEQUEÑO). • MATERIA. Diferencias significativas (TE MEDIO). • MÓDULO. El alumnado del módulo de formación básica mayor promedio que el del didáctico disciplinar (TE PEQUEÑO). • COMPETENCIA DOCENTE PERCIBIDA. Correlación positiva y pequeña. • COMPETENCIA DOCENTE PERCIBIDA. Diferencias significativas entre los tres grupos, donde los mayores promedios se concentran en el grupo de mayor competencia (TE MEDIO).
FACTOR 4	IMP.	• SEXO. Las mujeres mayor promedio (TE PEQUEÑO). • EDAD. Correlación positiva y muy baja. • GRUPOS EDAD. El grupo menor, menor promedio que los dos grupos (TE PEQUEÑO). • CURSO. El alumnado de 3º mayor promedio que el de 1º (TE PEQUEÑO). • UNIDAD VALORACIÓN. En módulos mayor promedio (TE MUY PEQUEÑO). • MATERIA. Diferencias significativas (TE PEQUEÑO). • MÓDULO. El alumnado de módulo de formación básica menor promedio que el de mención (TE PEQUEÑO). • EXPECTATIVAS. Quienes esperan aprobar menor promedio que los que manifiestan dudas (TE PEQUEÑO). • MODALIDAD. Mayor promedio en la modalidad completa (TE PEQUEÑO).
	REA.	• DEDICACIÓN FORMACIÓN. El alumnado que trabaja sin contrato menor promedio que el contratado (TE MUY PEQUEÑO). • GRADO. El alumnado de Educación Infantil menores promedios (TE PEQUEÑO). • CURSO. El alumnado de 2º mayor promedio que el de 1º y 3º (TE PEQUEÑO). • CENTRO. El alumnado de UD mayor promedio (TE MUY PEQUEÑO). • UNIDAD VALORACIÓN. En módulos menor promedio (TE MUY PEQUEÑO). • MATERIA. Diferencias significativas (TE GRANDE). • MÓDULO. El alumnado de formación básica menor promedio que el de mención, y el del didáctico disciplinar menor promedio que el de los tres restantes (TE PEQUEÑO). • RESULTADOS PARCIALES. Quienes disponen de resultados mayor promedio (TE PEQUEÑO). • COMPETENCIA DOCENTE PERCIBIDA. Correlación positiva y pequeña. • COMPETENCIA DOCENTE PERCIBIDA. Diferencias significativas entre los tres grupos, donde los mayores promedios se concentran en el grupo de mayor competencia (TE PEQUEÑO). • MODALIDAD. Mayor promedio en la modalidad abreviada (TE PEQUEÑO).

FACTOR	Sub.	CONCLUSIÓN
FACTOR 5	IMP.	• SEXO. Las mujeres mayor promedio. (TE. MUY PEQUEÑO) • EDAD. Correlación positiva y muy baja. • GRUPOS EDAD. El grupo de menor edad menor promedio que el de 20 años (TE PEQUEÑO). • CURSO. El alumnado de 3º mayor promedio que el de 1º (TE PEQUEÑO). • MATERIA. Diferencias significativas (TE PEQUEÑO). • MÓDULO. El alumnado de formación básica menor promedio que el del didáctico disciplinar (TE PEQUEÑO).
	REA.	• UNIDAD VALORACIÓN. En módulos menor promedio (TE PEQUEÑO). • MATERIA. Diferencias significativas (TE GRANDE). • MÓDULO. Varias diferencias: el promedio en el módulo didáctico disciplinar es inferior que el de formación básica y el de mención; por otro lado, el promedio registrado en DECA es menor que los módulos de formación básica y mención (TE PEQUEÑO). • RESULTADOS PARCIALES. Quienes disponen de resultados mayor promedio (TE PEQUEÑO). • EXPECTATIVAS. Quienes esperan aprobar mayor promedio que los que manifiestan dudas (TE PEQUEÑO). • COMPETENCIA DOCENTE PERCIBIDA. Correlación positiva y moderada. • COMPETENCIA DOCENTE PERCIBIDA. Diferencias significativas entre los tres grupos, donde los mayores promedios se concentran en el grupo de mayor competencia (TE GRANDE). • MODALIDAD. Mayor promedio en la modalidad abreviada (TE PEQUEÑO).
FACTOR 6	IMP.	• SEXO. Las mujeres mayor promedio (TE PEQUEÑO). • MATERIA. Diferencias significativas (TE PEQUEÑO).
	REA.	• UNIDAD VALORACIÓN. En módulos menor promedio (TE MUY PEQUEÑO). • MATERIA. Diferencias significativas (TE GRANDE). • MÓDULO. El alumnado de formación básica mayor promedio que el del módulo didáctico-disciplinar y DECA (TE PEQUEÑO). • EXPECTATIVAS. Quienes esperan aprobar mayor promedio que los que manifiestan dudas (TE PEQUEÑO). • COMPETENCIA DOCENTE PERCIBIDA. Correlación positiva y pequeña. • COMPETENCIA DOCENTE PERCIBIDA. Diferencias significativas entre los tres grupos, donde los mayores promedios se concentran en el grupo de mayor competencia (TE GRANDE). • MODALIDAD. Mayor promedio en la modalidad abreviada (TE PEQUEÑO).

Nota: Los factores se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. La letra cursiva sugiere una lectura cautelosa de los resultados asociados al factor sexto, por los indicios de ambigüedad evidenciados en su estructuración interna.

El recorrido vertical de la tabla 115 evidencia que las diferencias significativas se producen en todos los factores asociados a la competencia evaluativa docente y, mayoritariamente, en la subescala de realización. Este resultado es coherente con lo hallado en el análisis descriptivo inicial de las valoraciones discentes (objetivo F1), donde

las estimaciones de cumplimiento mostraban una mayor dispersión que las relativas a la importancia.

Para favorecer la exploración de las conclusiones anticipadas, los próximos apartados agrupan los resultados constatados en función de la naturaleza de las variables en las que se originan:

- En un primer bloque se presentan los resultados relativos a las variables socio-demográficas y profesionales (sexo, edad y dedicación a la formación).
- Las variables académico-institucionales (grado, curso, centro, módulo, materia y unidad de evaluación) conforman el segundo bloque del análisis.
- En el tercer bloque se exponen los resultados asociados a aquellas variables independientes que pueden representar una influencia importante en la valoración de la competencia evaluativa (experiencia previa con el profesorado, resultados parciales de aprendizaje recibidos, competencia docente percibida, expectativas académicas o resultado esperado y modalidad de respuesta).

3.2.1 Variables socio-demográficas y profesionales.

Según los resultados constatados, la valoración de la competencia evaluativa docente efectuada en la muestra discente difiere en función de las principales variables socio-demográficas y profesionales consideradas en el estudio. La comparación de los promedios de respuesta en función del **sexo** da lugar a varios resultados significativos, aunque con un tamaño de efecto pequeño o muy pequeño ($\eta^2 < 0,06$) (ver tabla 116).

Tabla 116
Valoración de la competencia evaluativa docente en la muestra discente en función del SEXO

Estadísticos de contraste (U de Mann Whitney)

Factores	U de Mann-Whitney	W de Wilcoxon	Z	Sig. asintót. (bilateral)	TE (η^2)
Total IMPORTANCIA	44074,000	74455,000	-4,690	,000	0,0313 PEQUEÑO
Total REALIZACIÓN	49408,000	153604,000	-2,606	,009	0,009 MUY PEQUEÑO
FACTOR 1 IMP.	46775,000	77156,000	-3,664	,000	0,0190 PEQUEÑO
FACTOR 1 REA.	48836,000	153032,000	-2,841	,004	0,002 MUY PEQUEÑO
FACTOR 2 IMP.	44952,000	75333,000	-4,391	,000	0,0274 PEQUEÑO
FACTOR 2 REA.	48439,500	152635,500	-2,994	,003	0,0128 PEQUEÑO
FACTOR 3 IMP.	50974,500	81355,500	-2,049	,040	0,006 MUY PEQUEÑO

Estadísticos de contraste (*U* de Mann Whitney)

Factores	<i>U</i> de Mann-Whitney	<i>W</i> de Wilcoxon	<i>Z</i>	Sig. asintót. (bilateral)	TE (η^2)
FACTOR 3 REA.	52573,500	154951,500	-1,203	,229	
FACTOR 4 IMP.	46402,000	76537,000	-3,846	,000	0,0210 PEQUEÑO
FACTOR 4 REA.	46250,500	131328,500	-,230	,818	
FACTOR 5 IMP.	50449,500	80584,500	-2,237	,025	0,007 MUY PEQUEÑO
FACTOR 5 REA.	52595,500	155880,500	-1,198	,231	
<i>FACTOR 6 IMP.</i>	<i>46612,000</i>	<i>76993,000</i>	<i>-3,762</i>	<i>,000</i>	<i>0,0203 PEQUEÑO</i>
<i>FACTOR 6 REA.</i>	<i>55278,000</i>	<i>159018,000</i>	<i>-,270</i>	<i>,787</i>	

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. Los resultados ligados a este último factor se representan con letra cursiva para indicar la lectura cautelosa que requiere su interpretación por indicios de ambigüedad en su estructuración interna.

Salvo en la realización atribuida a los cuatro últimos factores, la diferencia de promedios advertida entre alumnas y alumnos resulta generalizable a la población con una probabilidad de error inferior al 5% (IC = 95%; $p < .05$). Como evidencian los datos resumidos en la tabla 117, los resultados se ajustan a un mismo patrón en todas ellas: mientras que las mujeres atribuyen mayores puntuaciones en la subescala de relevancia, los hombres estiman mayores niveles de realización.

Tabla 117

Diferencias significativas en las valoraciones de la muestra discente en función del SEXO: estadísticos de grupo

Factores	Sexo	<i>n</i>	<i>M</i>	<i>s</i>	<i>SEM</i>
Total IMPORTANCIA	Hombre	246	4,2085	,35294	,02250
	Mujer	456	4,3415	,38185	,01788
Total REALIZACIÓN	Hombre	246	2,8840	,53447	,03408
	Mujer	456	2,7842	,65810	,03082
F1. Participación IMP.	Hombre	246	3,9909	,62325	,03974
	Mujer	456	4,1513	,66019	,03092
F1. Participación REA.	Hombre	246	2,3381	,88033	,05613
	Mujer	456	2,1656	,98895	,04631
F2. Planificación IMP.	Hombre	246	4,1878	,45357	,02892
	Mujer	456	4,3430	,43140	,02020
F2. Planificación REA.	Hombre	246	3,1260	,66435	,04236
	Mujer	456	2,9908	,77104	,03611

Factores	Sexo	n	M	s	SEM
F3. Ajustes evaluación IMP.	Hombre	246	4,3638	,54718	,03489
	Mujer	456	4,4346	,58973	,02762
F4. Regulación IMP.	Hombre	245	4,4367	,54590	,03488
	Mujer	456	4,5811	,51616	,02417
F5. Adecuación y... IMP.	Hombre	245	4,3673	,60544	,03868
	Mujer	456	4,4232	,72750	,03407
<i>F6. Control condiciones IMP.</i>	Hombre	246	4,1463	,53201	,03392
	Mujer	456	4,2844	,58505	,02740

La Rho de Spearman revela correlaciones significativas ($p < .05$) o muy significativas ($p < .01$) entre la **edad** del alumnado y varios de los factores asociados a la competencia evaluativa (ver tabla 118). La magnitud de las asociaciones es muy pequeña en todo caso ($r_s \leq .20$) y se produce con diferente signo en la subescala de importancia y en la de realización. La atribución de relevancia total y la referida a los Factores 2, 4 y 5 resulta mayor, cuanto mayor es la edad del alumnado; por contra, la relación entre la edad y la realización percibida es negativa en las correlaciones significativas detectadas.

Tabla 118

Correlación de la edad del alumnado y la valoración de la competencia evaluativa docente

COMPETENCIA DOCENTE	RHO de SPEARMAN		
	Coeficiente correlación	Sig. (bilateral)	N
TOTAL IMPORTANCIA	,102	,007	698
TOTAL REALIZACIÓN	-,090	,017	698
F1. Participación en la evaluación IMP.	,002	,953	698
F1. Participación en la evaluación REA.	-,139	,000	698
F2. Planificación de la evaluación IMP.	,151	,000	698
F2. Planificación de la evaluación REA.	-,107	,005	698
F3. Ajustes en la evaluación IMP.	,046	,228	698
F3. Ajustes en la evaluación REA.	-,045	,241	694
F4. Regulación del aprendizaje IMP.	,138	,000	697
F4. Regulación del aprendizaje REA.	-,035	,378	635
F5. Adecuación y precisión IMP.	,096	,012	697
F5. Adecuación y precisión REA.	-,045	,232	695
<i>F6. Control condiciones de evaluación IMP.</i>	-,009	,818	698
<i>F6. Control condiciones de evaluación REA.</i>	,012	,744	697

Nota: El sombreado gris indica las valoraciones en las que se producen correlaciones significativas ($p < .05$) y el azul aquellas otras en las que la asociación detectada resulta muy significativa ($p < .01$). Los resultados referidos al factor sexto se representan con letra cursiva para indicar la cautela que requiere su interpretación.

Para profundizar en esta tendencia de relación, en una segunda fase de análisis se ha procedido a la comparación de los promedios de respuesta registrados en cuatro grupos equiparables de edad dentro de la muestra de estudiantes (18 años y menores, 19 años, 20 años y 21 años y mayores). El contraste por medio de la H de Kruskal Wallis ha revelado seis diferencias estadísticamente significativas en este caso ($p < .05$), todas ellas de tamaño pequeño ($\eta^2 \leq 0,06$) (ver tabla 119):

Tabla 119

Valoración de la competencia evaluativa docente en la muestra discente en función de la EDAD

Estadísticos de contraste (H Kruskal-Wallis)

Factores	Chi-cuadrado	gl	Sig. asintót.	TE (η^2)
Total IMPORTANCIA	11,814	3	,008	0,0170 PEQUEÑO
Total REALIZACIÓN	6,216	3	,102	
FACTOR 1 IMP.	2,666	3	,446	
FACTOR 1 REA.	15,134	3	,002	0,0218 PEQUEÑO
FACTOR 2 IMP.	15,744	3	,001	0,0227 PEQUEÑO
FACTOR 2 REA.	11,492	3	,009	0,0165 PEQUEÑO
FACTOR 3 IMP.	6,653	3	,084	
FACTOR 3 REA.	2,529	3	,470	
FACTOR 4 IMP.	15,161	3	,002	0,0218 PEQUEÑO
FACTOR 4 REA.	5,650	3	,130	
FACTOR 5 IMP.	13,735	3	,003	0,0198 PEQUEÑO
FACTOR 5 REA.	2,525	3	,471	
<i>FACTOR 6 IMP.</i>	<i>7,482</i>	<i>3</i>	<i>,058</i>	
<i>FACTOR 6 REA.</i>	<i>4,735</i>	<i>3</i>	<i>,192</i>	

Nota: El sombreado gris indica las valoraciones en las que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. Los resultados referidos a este último factor se representan con letra cursiva para indicar la cautela que requiere su interpretación.

La opción paramétrica del análisis de varianza (ANOVA) ha permitido concluir que el grupo de estudiantes de 19 años no participa de ninguna de las diferencias significativas determinadas ($p > .05$) (ver tabla 120). Confirmando la relación de asociación detectada por medio de la Rho de Spearman (tabla 118), la estimación de cumplimiento al valorar la Planificación de la evaluación (F2) y la Participación discente en la misma (F1) es inferior en el alumnado de mayor edad. En ambos casos los promedios superiores los registra el alumnado más joven de la muestra (17 y 18 años); en el factor primero las diferencias se

producen con respecto a los dos grupos de mayor edad (20 años y 21 años y más), mientras que en el segundo sólo en contraste con el de los sujetos de 20 años.

La correlación positiva entre la edad del alumnado y las valoraciones de relevancia también queda reafirmada en el contraste de promedios para los Factores 2 (Planificación de la evaluación), 4 (Regulación del aprendizaje) y 5 (Adecuación y precisión de la evaluación). Aunque las diferencias involucran a diferentes grupos de edad, en los tres casos el alumnado de menor edad destaca como el que menores puntuaciones otorga.

Tabla 120

Diferencias significativas en la valoración de la competencia evaluativa docente en función de la EDAD del alumnado: estadísticos de grupo y contraste post hoc (Bonferroni y Scheffé)

Diferencias significativas en función de la EDAD		<i>n</i>	<i>M</i>	<i>s</i>	Sig. 17-18 años	Sig. 19 años	Sig. 20 años	Sig. ≥ 21 años
TOTAL IMPORTANCIA	17-18 años	194	4,2347	,38066	-	.773	.009	.384
	19 años	162	4,2769	,35285	.773	-	.174	.933
	20 años	178	4,3679	,37253	.009		-	
	21 años y más	164	4,3043	,39445	.384	.933	.174	-
FACTOR 1 REALIZACIÓN	17-18 años	194	2,4291	,94427	-	.233	.030	.008
	19 años	162	2,2207	1,0009	.233	-	.873	.633
	20 años	178	2,1348	,91291	.030	.873	-	.969
	21 años y más	164	2,0833	,92626	.008	.633	.969	-
FACTOR 2 IMPORTANCIA	17-18 años	194	4,2134	,44422	-	.908	.030	.018
	19 años	162	4,2481	,47224	.908	-	.207	.143
	20 años	178	4,3506	,41364	.030	.207	-	.996
	21 años y más	164	4,3622	,43501	.018	.143	.996	-
FACTOR 2 REALIZACIÓN	17-18 años	194	3,1438	,78191	-	.988	.032	.132
	19 años	162	3,1160	,71078	.988	-	.119	.293
	20 años	178	2,9233	,72416	.032	.119	-	.976
	21 años y más	164	2,9595	,69896	.132	.293	.976	-
FACTOR 4 IMPORTANCIA	17-18 años	194	4,4364	,53407	-	.942	.034	.045
	19 años	162	4,5165	,52284	.942	-	1	1
	20 años	177	4,5895	,50221	.034	1	-	1
	21 años y más	164	4,5874	,56344	.045	1	1	-
FACTOR 5 IMPORTANCIA	17-18 años	194	4,2423	,81949	-	.051	.000	.176
	19 años	162	4,4444	,62306	.051	-	.416	.199
	20 años	178	4,5534	,53278	.000	.416	-	.209
	21 años y más	163	4,4141	,68627	.176	.199	.209	-

Nota: Las celdas sombreadas señalan los grupos entre los que tiene lugar la diferencia de promedios dentro del intervalo de confianza establecido ($p < .05$).

La **dedicación del alumnado a su formación universitaria** cierra el análisis de este primer bloque de variables. Sujeta a tres posibles valores dentro de la muestra (dedicación exclusiva a la formación, combinación de la formación con un trabajo regulado o combinación con una actividad remunerada sin regulación contractual), la comparación de promedios en función de la misma revela diferencias estadísticamente significativas dentro del intervalo de confianza fijado (IC = 95%). Resumidas en la tabla 121, el tamaño del efecto estimado para las mismas es pequeño ($0,01 < \eta^2 < 0,06$) o muy pequeño ($\eta^2 < 0,01$).

Tabla 121

Valoración de la competencia evaluativa docente en la muestra discente en función de la DEDICACIÓN A LA FORMACIÓN UNIVERSITARIA

Estadísticos de contraste (H Kruskal-Wallis)				
Factores	Chi-cuadrado	gl	Sig. asintót.	TE (η^2)
Total IMPORTANCIA	8,335	2	,015	0,0119 PEQUEÑO
Total REALIZACIÓN	1,265	2	,531	
FACTOR 1 IMP.	3,664	2	,160	
FACTOR 1 REA.	1,525	2	,467	
FACTOR 2 IMP.	15,278	2	,000	0,0218 PEQUEÑO
FACTOR 2 REA.	,529	2	,768	
FACTOR 3 IMP.	6,262	2	,044	0,0089 MUY PEQUEÑO
FACTOR 3 REA.	,686	2	,710	
FACTOR 4 IMP.	5,230	2	,073	
FACTOR 4 REA.	6,237	2	,044	0,0089 MUY PEQUEÑO
FACTOR 5 IMP.	3,524	2	,172	
FACTOR 5 REA.	1,219	2	,544	
<i>FACTOR 6 IMP.</i>	<i>,603</i>	<i>2</i>	<i>,740</i>	
<i>FACTOR 6 REA.</i>	<i>,143</i>	<i>2</i>	<i>,931</i>	

Nota: El sombreado gris indica las valoraciones en las que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. Los resultados referidos a este último factor se representan con letra cursiva para indicar la cautela que requiere su interpretación.

Como indica la tabla 122, el contraste *post hoc* efectuado por medio del ANOVA, permite detectar que, en las tres valoraciones de importancia, las diferencias significativas se producen entre el alumnado con dedicación exclusiva a la formación y el que además trabaja sin contrato. En el promedio total de la subescala ($p = .014$) o a la hora de valorar la relevancia de los factores primero y tercero sobre Planificación evaluativa ($p = .014$) y los Ajustes realizados en la evaluación ($p = .020$), el alumnado que no se dedica en exclusiva a

sus estudios manifiesta puntuaciones sensiblemente superiores en comparación con el que sí lo hace.

Tabla 122

Diferencias significativas en las valoraciones de la muestra docente en función de la DEDICACIÓN A LA FORMACIÓN: estadísticos de grupo y contraste post hoc (Bonferroni, DMS y Scheffé)

Diferencias significativas en función de la DEDICACIÓN		<i>n</i>	<i>M</i>	<i>s</i>	Sig. Exclusiva	Sig. Contrato	Sig. Sin contrato
TOTAL IMPORTANCIA	Exclusiva	399	4,2562	,38242	-	.310	.014
	Contrato	61	4,3354	,38236	.310	-	.980
	Sin contrato	242	4,3462	,36270	.014	.980	-
FACTOR 2 IMPORTANCIA	Exclusiva	399	4,2361	,43710	-	.069	.005
	Contrato	61	4,3770	,47763	.069	-	.939
	Sin contrato	242	4,3545	,44077	.014	.939	-
FACTOR 3 IMPORTANCIA	Exclusiva	399	4,3609	,60284	-	.062	.020
	Contrato	61	4,5082	,51498	.062	-	.640
	Sin contrato	242	4,4697	,53744	.020	.640	-
FACTOR 4 REALIZACIÓN	Exclusiva	358	2,3580	,95938	-	.174	.089
	Contrato	58	2,5402	,90014	.174	-	.022
	Sin contrato	224	2,2210	,93453	.089	.022	-

Nota: Las celdas sombreadas señalan los grupos entre los que tiene lugar la diferencia de promedios dentro del intervalo de confianza establecido ($p < .05$).

En la realización que percibe el colectivo discente de la Regulación del aprendizaje por medio de la evaluación (Factor 4), se produce una diferencia de efecto muy pequeño entre los dos grupos que no se dedican en exclusiva a su formación universitaria. Aquel grupo en el que la actividad adicional se lleva a cabo sin contrato percibe un cumplimiento levemente inferior en este aspecto ($p = .022$).

3.2.2 Variables académico-institucionales.

Este segundo bloque incorpora seis variables de identificación relacionadas con factores institucionales y académicos vinculados al alumnado. La primera de ellas es el **grado o titulación**, que divide la muestra en dos grupos: el alumnado matriculado en Educación Infantil y aquel otro que cursa sus estudios en el Educación Primaria.

Tabla 123

*Valoración de la competencia evaluativa docente en la muestra discente en función del GRADO***Estadísticos de contraste** (*U* de Mann Whitney)

Factores	<i>U</i> de Mann-Whitney	<i>W</i> de Wilcoxon	<i>Z</i>	Sig. asintót. (bilateral)	TE (η^2)
Total IMPORTANCIA	37827,000	184438,000	-2,754	,006	0,0108 PEQUEÑO
Total REALIZACIÓN	37393,500	50759,500	-2,943	,003	0,0123 PEQUEÑO
FACTOR 1 IMP.	39863,000	186474,000	-1,874	,061	
FACTOR 1 REA.	36254,000	49620,000	-3,458	,001	0,0170 PEQUEÑO
FACTOR 2 IMP.	38914,500	185525,500	-2,299	,022	0,007 MUY PEQUEÑO
FACTOR 2 REA.	36558,000	49924,000	-3,321	,001	0,0157 PEQUEÑO
FACTOR 3 IMP.	38146,000	184757,000	-2,684	,007	0,0103 PEQUEÑO
FACTOR 3 REA.	40237,500	185228,500	-1,500	,134	
FACTOR 4 IMP.	39858,500	185928,500	-1,899	,058	
FACTOR 4 REA.	29364,000	42730,000	-4,726	,000	0,0318 PEQUEÑO
FACTOR 5 IMP.	42055,000	188666,000	-,823	,410	
FACTOR 5 REA.	40543,000	53423,000	-1,229	,219	
<i>FACTOR 6 IMP.</i>	<i>40494,000</i>	<i>187105,000</i>	<i>-1,609</i>	<i>,108</i>	
<i>FACTOR 6 REA.</i>	<i>42118,000</i>	<i>188188,000</i>	<i>-,838</i>	<i>,402</i>	

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. Los resultados ligados a este último factor se representan con letra cursiva para indicar la lectura cautelosa que requiere su interpretación por los indicios de ambigüedad en su estructuración interna.

Comparadas las valoraciones realizadas por el alumnado de ambas titulaciones a través de la *U* de Mann Whitney, se detectan siete diferencias estadísticamente significativas ($p \leq .022$) de efecto pequeño o muy pequeño (ver tabla 123). Los datos contenidos en la tabla 124 sugieren para las mismas un mismo patrón respuesta: mientras que el alumnado del Grado de Educación Infantil ofrece mayores puntuaciones al atribuir niveles de relevancia, queda por debajo del de Educación Primaria al estimar grados de realización.

Tabla 124

Diferencias significativas en las valoraciones de la muestra discente en función del GRADO: estadísticos de grupo

Factores	Grado	n	M	s	SEM
Total IMPORTANCIA	E. Primaria	541	4,2765	,36507	,01570
	E. Infantil	163	4,3510	,41134	,03222

Factores	Grado	n	M	s	SEM
Total REALIZACIÓN.	E. Primaria	541	2,8581	,62909	,02705
	E. Infantil	163	2,6938	,56663	,04438
F1. Participación REA.	E. Primaria	541	2,2930	,96699	,04157
	E. Infantil	163	1,9918	,87920	,06886
F2. Planificación IMP.	E. Primaria	541	4,2680	,45112	,01940
	E. Infantil	163	4,3595	,41761	,03271
F2. Planificación REA.	E. Primaria	541	3,0832	,72411	,03113
	E. Infantil	163	2,8908	,76397	,05984
F3. Ajustes evaluación IMP.	E. Primaria	541	4,3814	,58024	,02495
	E. Infantil	163	4,5051	,55104	,04316
F4. Regulación REA.	E. Primaria	478	2,4282	,95579	,04372
	E. Infantil	163	2,0225	,86156	,06748

El **curso** aparece asociado a una gran cantidad de resultados generalizables a la población ($p < .05$), aunque con una magnitud de efecto pequeña ($0,01 < \eta^2 < 0,06$) (ver tabla 125):

Tabla 125

Valoración de la competencia evaluativa docente en la muestra discente en función del CURSO

Estadísticos de contraste (H Kruskal-Wallis)

Factores	Chi-cuadrado	gl	Sig. asintót.	TE (η^2)
Total IMPORTANCIA	15,701	2	,000	0,0224 PEQUEÑO
Total REALIZACIÓN	22,495	2	,000	0,0320 PEQUEÑO
FACTOR 1 IMP.	7,056	2	,029	0,0101 PEQUEÑO
FACTOR 1 REA.	36,702	2	,000	0,0523 PEQUEÑO
FACTOR 2 IMP.	17,498	2	,000	0,0249 PEQUEÑO
FACTOR 2 REA.	23,438	2	,000	0,0334 PEQUEÑO
FACTOR 3 IMP.	1,458	2	,483	
FACTOR 3 REA.	12,809	2	,002	0,0180 PEQUEÑO
FACTOR 4 IMP.	23,868	2	,000	0,0340 PEQUEÑO
FACTOR 4 REA.	24,395	2	,000	0,0348 PEQUEÑO
FACTOR 5 IMP.	14,883	2	,001	0,0212 PEQUEÑO
FACTOR 5 REA.	2,354	2	,308	
FACTOR 6 IMP.	1,340	2	,512	
FACTOR 6 REA.	5,022	2	,081	

Nota: El sombreado gris indica las valoraciones en las que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los resultados referidos al factor sexto se representan con letra cursiva para indicar la cautela que requiere su interpretación.

La tabla 126 recopila las conclusiones de las exploraciones *post hoc* efectuadas por medio del ANOVA para determinar entre qué grupos se producen las diferencias. Su análisis sugiere la incidencia de distintos patrones de relación y entre diferentes cursos:

- En la subescala de importancia, el alumnado del tercer curso ofrece puntuaciones significativamente superiores a las de las/os estudiantes de primero ($p = .000$). Esta tendencia se refleja exactamente igual en la valoración del nivel de relevancia de los factores quinto (Adecuación y precisión), cuarto (Regulación del aprendizaje) y segundo (Planificación de la evaluación). Al considerar la importancia de la Participación discente en la evaluación (Factor 1), el alumnado de tercero sigue mostrando las puntuaciones más elevadas, aunque respecto a las que se registran el segundo año de la titulación ($p = .031$).
- En la subescala de realización las diferencias estadísticamente significativas ($p < .05$) se producen entre la percepción comunicada por estudiantes de primero y la referida en los dos cursos superiores. En todas ellas los mayores promedios los ofrece el alumnado de primero; así se constata también al valorar el cumplimiento percibido en los *Ajustes del sistema evaluativo* (Factor 3), aunque en este caso la discrepancia de puntuaciones sólo es significativa al comparar los dos grupos extremos (1º y 3º) ($p = .005$).

Tabla 126

Diferencias significativas en las valoraciones de la muestra docente en función del CURSO: estadísticos de grupo y contraste post hoc (Bonferroni, DMS y Scheffé)

Diferencias significativas en función de la CURSO		<i>n</i>	<i>M</i>	<i>s</i>	Sig. 1º curso	Sig. 2º curso	Sig. 3º curso
TOTAL IMPORTANCIA	1º Curso	358	4,2463	,37693	-	.573	.000
	2º Curso	123	4,2875	,38468	.573	-	.125
	3º Curso	223	4,3733	,36192	.000	.125	-
TOTAL REALIZACIÓN	1º Curso	358	2,9320	,64524	-	.000	.000
	2º Curso	123	2,7070	,47270	.000	-	1
	3º Curso	223	2,7027	,61542	.000	1	-
FACTOR 1 IMPORTANCIA	1º Curso	358	4,0803	,65999	-	.289	.306
	2º Curso	123	3,9715	,68667	.289	-	.031
	3º Curso	223	4,1670	,64407	.306	.031	-
FACTOR 1 REALIZACIÓN	1º Curso	358	2,4372	,97620	-	.000	.000
	2º Curso	123	1,9492	,79248	.000	-	.071
	3º Curso	223	2,0310	,92715	.000	.071	-

Diferencias significativas en función de la CURSO		<i>n</i>	<i>M</i>	<i>s</i>	Sig. 1º curso	Sig. 2º curso	Sig. 3º curso
FACTOR 2 IMPORTANCIA	1º Curso	358	4,2218	,46847	-	.074	.000
	2º Curso	123	4,3268	,41470	.074	-	.601
	3º Curso	223	4,3767	,40470	.000	.601	-
FACTOR 2 REALIZACIÓN	1º Curso	358	3,1534	,76571	-	.424	.000
	2º Curso	123	3,0565	,63652	.424	-	.014
	3º Curso	223	2,8446	,70533	.000	.014	-
FACTOR 3 REALIZACIÓN	1º Curso	358	3,9120	,99631	-	.430	.005
	2º Curso	123	3,7737	1,0999	.430	-	.406
	3º Curso	219	3,6279	1,0045	.005	.406	-
FACTOR 4 IMPORTANCIA	1º Curso	358	4,4525	,55627	-	.481	.000
	2º Curso	122	4,5246	,52121	.481	-	.067
	3º Curso	223	4,6547	,47337	.000	.067	-
FACTOR 4 REALIZACIÓN	1º Curso	313	2,4318	,87765	-	.000	.957
	2º Curso	123	1,9431	,80719	.000	-	.000
	3º Curso	205	2,3911	1,07053	.957	.000	-
FACTOR 5 IMPORTANCIA	1º Curso	358	4,3059	,74709	-	.139	.000
	2º Curso	123	4,4472	,65853	.139	-	.474
	3º Curso	222	4,5405	,56734	.000	.474	-

Nota: Las celdas sombreadas señalan los grupos entre los que tiene lugar la diferencia de promedios dentro del intervalo de confianza establecido ($p < .05$).

Comparadas las valoraciones del colectivo discente en función del **centro** de procedencia (UD o BAM, su centro adscrito) se ponen de manifiesto tres nuevas diferencias significativas dentro de la muestra, dos ligadas a la subescala de realización y una tercera asociada a la de relevancia (ver tabla 127).

Tabla 127

Valoración de la competencia evaluativa docente en la muestra discente en función del CENTRO

Estadísticos de contraste (U de Mann Whitney)

Factores	U de Mann-Whitney	W de Wilcoxon	Z	Sig. asintót. (bilateral)	TE (η^2)
Total IMPORTANCIA	51109,500	72845,500	-,193	,847	
Total REALIZACIÓN	51069,500	72805,500	-,209	,834	
FACTOR 1 IMP.	46084,500	169340,500	-2,253	.024	0,0072 MUY PEQUEÑO
FACTOR 1 REA.	48533,500	171789,500	-1,245	,213	
FACTOR 2 IMP.	47670,500	69406,500	-1,607	,108	
FACTOR 2 REA.	46308,000	68044,000	-2,150	.032	0,0066 MUY PEQUEÑO
FACTOR 3 IMP.	47779,500	69515,500	-1,588	,112	

Estadísticos de contraste (*U* de Mann Whitney)

Factores	<i>U</i> de Mann-Whitney	<i>W</i> de Wilcoxon	<i>Z</i>	Sig. asintót. (bilateral)	TE (η^2)
FACTOR 3 REA.	48641,500	70377,500	-1,047	,295	
FACTOR 4 IMP.	51212,500	173972,500	-,113	,910	
FACTOR 4 REA.	36632,500	144048,500	-2,191	,028	0,0075 MUY PEQUEÑO
FACTOR 5 IMP.	48236,000	170996,000	-1,395	,163	
FACTOR 5 REA.	48169,500	169940,500	-1,280	,200	
<i>FACTOR 6 IMP.</i>	<i>51258,000</i>	<i>72994,000</i>	<i>-,135</i>	<i>,893</i>	
<i>FACTOR 6 REA.</i>	<i>46715,500</i>	<i>68451,500</i>	<i>-1,950</i>	<i>,051</i>	

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. Los resultados ligados a este último factor se representan con letra cursiva para indicar la lectura cautelosa que requiere su interpretación por los indicios de ambigüedad en su estructuración interna.

Aunque se encuentran dentro del intervalo de confianza establecido, el tamaño del efecto estimado para estas diferencias es muy pequeño ($\eta^2 \leq 0,01$). Las valoraciones de las/os estudiantes del Campus bilbaíno de la UD son superiores al juzgar la importancia de la Participación discente en la evaluación (Factor 1) y la realización del factor cuarto (Regulación del aprendizaje) ($p = .024$ y $p = .028$, respectivamente); en el caso de la Planificación de la evaluación (Factor 2) es el alumnado de BAM el que percibe mayor grado de cumplimiento ($p = .028$). Así se expone en la tabla 128, que sintetiza los estadísticos referidos a dichas valoraciones:

Tabla 128

Diferencias significativas en las valoraciones de la muestra discente en función del CENTRO: estadísticos de grupo

Factores	Centro	<i>n</i>	<i>M</i>	<i>s</i>	<i>SEM</i>
F1. Participación IMP.	BAM	496	4,0625	,64305	,02887
	UD	208	4,1514	,70318	,04876
F2. Planificación REA.	BAM	496	3,0801	,75056	,03370
	UD	208	2,9397	,69708	,04833
F4. Regulación REA.	BAM	463	2,2729	,94297	,04382
	UD	178	2,4607	,95275	,07141

De acuerdo al procedimiento expuesto en el capítulo anterior, el alumnado de la muestra ha participado en la investigación tomando como referencia el sistema evaluativo empleado por un/a profesor/a en un módulo o asignatura determinada. Dependiendo de si

la valoración se ha efectuado para **unidades de docencia** de 3 o 6 ECTS (módulos específicos o asignaturas completas), las respuestas del alumnado difieren dentro del intervalo de confianza establecido en determinados casos:

Tabla 129

Valoración de la competencia evaluativa docente en la muestra discente en función de la UNIDAD de REFERENCIA (módulo o asignatura)

Estadísticos de contraste (*U* de Mann Whitney)

Factores	<i>U</i> de Mann-Whitney	<i>W</i> de Wilcoxon	<i>Z</i>	Sig. asintót. (bilateral)	TE (η^2)
Total IMPORTANCIA	32792,500	198392,500	-2,059	,040	0,0060 MUY PEQUEÑO
Total REALIZACIÓN	23473,000	31858,000	-6,522	,000	0,065 MEDIO
FACTOR 1 IMP.	36807,500	202407,500	-,135	,892	
FACTOR 1 REA.	24103,000	32488,000	-6,247	,000	0,055 PEQUEÑO
FACTOR 2 IMP.	33034,000	198634,000	-1,963	,050	
FACTOR 2 REA.	24711,000	33096,000	-5,949	,000	0,0504 PEQUEÑO
FACTOR 3 IMP.	35600,500	201200,500	-,732	,464	
FACTOR 3 REA.	26278,500	34279,500	-4,873	,000	0,0340 PEQUEÑO
FACTOR 4 IMP.	32323,500	197348,500	-2,343	,019	0,0078 MUY PEQUEÑO
FACTOR 4 REA.	28402,500	36403,500	-2,182	,029	0,0074 MUY PEQUEÑO
FACTOR 5 IMP.	34965,000	200565,000	-,933	,351	
FACTOR 5 REA.	28054,000	36310,000	-4,206	,000	0,0253 PEQUEÑO
<i>FACTOR 6 IMP.</i>	<i>34704,500</i>	<i>200304,500</i>	<i>-1,162</i>	<i>,245</i>	
<i>FACTOR 6 REA.</i>	<i>29635,500</i>	<i>37891,500</i>	<i>-3,469</i>	<i>,001</i>	<i>0,0172 PEQUEÑO</i>

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. Los resultados ligados a este último factor se representan con letra cursiva para indicar la lectura cautelosa que requiere su interpretación por los indicios de ambigüedad en su estructuración interna.

Comparados los promedios de respuesta de ambos subgrupos por medio de la *U* de Mann Whitney, se producen las ocho diferencias significativas ($p < .05$) destacadas en la tabla 129. Como en las variables anteriores, el tamaño del efecto atribuido a las mismas es pequeño o muy pequeño salvo en la valoración total de realización, donde la puntuación eta cuadrada indica un efecto de tamaño medio.

Tabla 130

Diferencias significativas en las valoraciones de la muestra discente en función de la UNIDAD de DOCENCIA valorada: estadísticos de grupo

Factores	Unidad	n	M	s	SEM
Total IMPORTANCIA	Asignatura	575	4,2802	,37831	,01578
	Módulo	129	4,3540	,36819	,03242
Total REALIZACIÓN.	Asignatura	575	2,8932	,62240	,02596
	Módulo	129	2,4941	,48381	,04260
F1. Participación REA.	Asignatura	575	2,3283	,96688	,04032
	Módulo	129	1,7552	,74118	,06526
F2. Planificación REA.	Asignatura	575	3,1169	,74955	,03126
	Módulo	129	2,6899	,56345	,04961
F3. Ajustes evaluación REA.	Asignatura	574	3,8815	1,0141	,04233
	Módulo	126	3,4220	,98741	,08797
F4. Regulación IMP.	Asignatura	574	4,5070	,54527	,02276
	Módulo	129	4,6279	,45763	,04029
F4. Regulación REA.	Asignatura	515	2,3676	,95444	,04206
	Módulo	126	2,1508	,90785	,08088
F5. Adecuación... REA.	Asignatura	573	3,1091	1,0151	,04241
	Módulo	128	2,6992	,97879	,08651
F6. Control evaluación REA.	Asignatura	575	3,4783	,89805	,03745
	Módulo	128	3,2083	,74682	,06601

Los datos estadísticos resumidos en la tabla 130 revelan que las diferencias constatadas se ajustan a un mismo patrón de respuesta: en las valoraciones de importancia las/os estudiantes de módulos producen mayores promedios de respuesta que el alumnado de asignaturas completas, mientras que al determinar niveles de cumplimiento percibido la tendencia se invierte. Así se refleja en la puntuación total de la subescala de realización, donde las valoraciones efectuadas para asignaturas completas superan en 0,4 puntos las producidas para las unidades de 3 ECTS ($\eta^2 = 0,065$). Atendiendo a las dificultades reconocidas por el colectivo docente para el ejercicio de las funciones evaluativas (objetivo E3), este resultado pudiera estar relacionado con la mayor limitación temporal aplicable a los módulos, donde el tiempo de trabajo disponible se reduce ostensiblemente.

En la formación básica de maestras/os de Educación Infantil y Primaria, el módulo o asignatura en la que se haya llevado a cabo la valoración se inscribe en una determinada **materia**, según las órdenes de referencia (ECI/3854/2007 y ECI/3857/2007, respectivamente). Comparando las respuestas discentes en función de las mismas, se aprecian diferencias significativas con un margen de error inferior al 5% en todos los factores y en ambas subescalas. Recogidas en la tabla 131, el tamaño del efecto estimado

para las mismas varía según el caso: en las valoraciones de realización resulta medio o grande, mientras que en los promedios de relevancia el efecto es pequeño, a excepción del total de la subescala, donde llega a ser medio.

Tabla 131

Valoración de la competencia evaluativa docente en la muestra discente en función de la MATERIA

Estadísticos de contraste (H Kruskal-Wallis)

Factores	Chi-cuadrado	gl	Sig. asintót.	TE (η^2)
Total IMPORTANCIA	49,963	15	,000	0,0711 MEDIO
Total REALIZACIÓN	187,657	15	,000	0,2669 GRANDE
FACTOR 1 IMP.	41,539	15	,000	0,0591 PEQUEÑO
FACTOR 1 REA.	111,803	15	,000	0,1590 GRANDE
FACTOR 2 IMP.	31,125	15	,008	0,0443 PEQUEÑO
FACTOR 2 REA.	207,196	15	,000	0,2947 GRANDE
FACTOR 3 IMP.	27,377	15	,026	0,0389 PEQUEÑO
FACTOR 3 REA.	80,571	15	,000	0,1146 MEDIO
FACTOR 4 IMP.	38,315	15	,001	0,0545 PEQUEÑO
FACTOR 4 REA.	170,135	14	,000	0,2420 GRANDE
FACTOR 5 IMP.	24,984	15	,050	0,0355 PEQUEÑO
FACTOR 5 REA.	114,029	15	,000	0,1622 GRANDE
<i>FACTOR 6 IMP.</i>	<i>25,795</i>	<i>15</i>	<i>,040</i>	<i>0,0367 PEQUEÑO</i>
<i>FACTOR 6 REA.</i>	<i>117,822</i>	<i>15</i>	<i>,000</i>	<i>0,1676 GRANDE</i>

Nota: El sombreado gris indica las valoraciones en las que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. Los resultados referidos a este último factor se representan con letra cursiva para indicar la cautela que requiere su interpretación.

Considerando las materias comunes y específicas que se proponen para cada titulación, esta variable suma 24 grupos o valores diferentes dentro de la muestra. Por ese motivo, en vez de realizar una exploración *post hoc* por la vía paramétrica, su análisis se ha efectuado indirectamente, por medio de otras variables independientes que la delimitan naturalmente. Tal es el caso del **módulo**, que agrupa las diversas materias en tres grupos diferentes reconocidos en los decretos reguladores para ambas titulaciones (formación básica, didáctico-disciplinar y mención); a ellos se suma un cuarto valor creado para responder a la particularidad de la oferta formativa institucional (Declaración Eclesiástica de Competencia Académica, DECA).

Analizadas por medio de la *H* de Kruskal Wallis, las respuestas del alumnado de los cuatro módulos evidencian diferencias dentro del nivel de confianza establecido (IC = 95%)

(ver tabla 132). En coherencia con los resultados hallados en la variable materia, se vuelven a constatar diferencias significativas en todas las valoraciones de la subescala de realización, aunque en este caso con un efecto de tamaño medio ($0,06 < \eta^2 < 0,14$). En la Regulación del aprendizaje por medio de la evaluación (Factor 4) y en la Adecuación y precisión evaluativa (Factor 5) la diferencia de promedios también es significativa al valorar la relevancia ($p = .011$ y $p = .045$, respectivamente):

Tabla 132

Valoración de la competencia evaluativa docente en la muestra discente en función del MÓDULO

Estadísticos de contraste (*H* Kruskal-Wallis)

Factores	Chi-cuadrado	gl	Sig. asintót.	TE (η^2)
Total IMPORTANCIA	6,528	3	,089	
Total REALIZACIÓN	37,397	3	,000	0,0532 PEQUEÑO
FACTOR 1 IMP.	2,715	3	,438	
FACTOR 1 REA.	32,125	3	,000	0,0457 PEQUEÑO
FACTOR 2 IMP.	7,536	3	,057	
FACTOR 2 REA.	27,068	3	,000	0,0385 PEQUEÑO
FACTOR 3 IMP.	3,206	3	,361	
FACTOR 3 REA.	17,775	3	,000	0,0253 PEQUEÑO
FACTOR 4 IMP.	11,213	3	,011	0,0160 PEQUEÑO
FACTOR 4 REA.	38,322	3	,000	0,0545 PEQUEÑO
FACTOR 5 IMP.	8,027	3	,045	0,0114 PEQUEÑO
FACTOR 5 REA.	12,405	3	,006	0,0176 PEQUEÑO
<i>FACTOR 6 IMP.</i>	<i>1,357</i>	<i>3</i>	<i>,716</i>	
<i>FACTOR 6 REA.</i>	<i>22,646</i>	<i>3</i>	<i>,000</i>	<i>0,0322 PEQUEÑO</i>

Nota: El sombreado gris indica las valoraciones en las que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. Los resultados referidos a este último factor se representan con letra cursiva para indicar la cautela que requiere su interpretación.

Como en casos anteriores, para indagar sobre la raíz de tales diferencias se ha recurrido a los análisis *post hoc* que ofrece la alternativa paramétrica del análisis de varianza (ANOVA). Como refleja la tabla 133, la discrepancia de percepciones no siempre se produce dentro de los mismos grupos.

Tabla 133

Diferencias significativas en la valoración de la competencia evaluativa docente en función del MÓDULO: estadísticos de grupo y contraste post hoc (T2 de Tamhane, T3 de Dunnet, C de Dunnet, Games-Howel, Bonferroni y Scheffé)

Diferencias significativas en función del MÓDULO		<i>n</i>	<i>M</i>	<i>s</i>	Sig. FB	Sig. DD	Sig. MENCIÓN	Sig. DECA
TOTAL REALIZACIÓN	FB	343	2,9438	,63555	-	.000	.998	.053
	DD	262	2,6465	,55019	.000	-	.014	1
	MENCIÓN	77	2,9057	,66367	.998	.014	-	.254
	DECA	22	2,6589	,44683	.053	1	.254	-
FACTOR 1 REALIZACIÓN	FB	343	2,4031	,98652	-	.000	.668	.000
	DD	262	2,0331	,89822	.000	-	.362	.047
	MENCIÓN	77	2,2435	,89404	.668	.362	-	.003
	DECA	22	1,6136	,63960	.000	.047	.003	-
FACTOR 2 REALIZACIÓN	FB	343	3,1687	,78083	-	.000	1	.080
	DD	262	2,8508	,63671	.000	-	.023	.999
	MENCIÓN	77	3,1409	,78885	1	.023	-	.328
	DECA	22	2,8909	,45660	.080	.999	.328	-
FACTOR 3 REALIZACIÓN	FB	343	3,9417	1,0179	-	.017	.230	.244
	DD	261	3,6737	1,0248	.017	-	1	.873
	MENCIÓN	74	3,6712	,94963	.230	1	-	.903
	DECA	22	3,4848	1,0970	.244	.873	.903	-
FACTOR 4 IMPORTANCIA	FB	343	4,4713	,54469	-	.203	.007	.751
	DD	261	4,5658	,55435	.203	-	.638	1
	MENCIÓN	77	4,6450	,38365	.007	.638	-	.995
	DECA	22	4,5909	,41060	.751	1	.995	-
FACTOR 4 REALIZACIÓN	FB	298	2,3641	,90199	-	.003	.101	.004
	DD	251	2,0976	,86506	.003	-	.000	.000
	MENCIÓN	70	2,7190	1,1500	.101	.000	-	.403
	DECA	22	3,1364	,90653	.004	.000	.403	-
FACTOR 5 IMPORTANCIA	FB	343	4,3324	,72872	-	.009	.265	.118
	DD	261	4,4789	,66255	.009	-	.551	.577
	MENCIÓN	77	4,4286	,58890	.265	.551	-	.399
	DECA	22	4,5682	,49511	.118	.577	.399	-
FACTOR 5 REALIZACIÓN	FB	341	3,1261	1,0052	-	.009	.700	.022
	DD	261	2,9080	,98017	.009	-	.042	.191
	MENCIÓN	77	3,1753	1,1348	.700	.042	-	.635
	DECA	22	2,6136	1,0903	.022	.191	.635	-

Diferencias significativas en función del MÓDULO		<i>n</i>	<i>M</i>	<i>s</i>	Sig. FB	Sig. DD	Sig. MENCIÓN	Sig. DECA
FACTOR 6 REALIZACIÓN	FB	343	3,5637	,90118	-	.001	.701	.004
	DD	262	3,2920	,85997	.001	-	.723	.250
	MENCIÓN	76	3,4276	,77523	.701	.723	-	.062
	DECA	22	2,9697	,68165	.004	.250	.062	-

Nota: La comparación se realiza entre cuatro módulos: Formación Básica (FB), Didáctico Disciplinar (DD), Mención y DECA. Las celdas sombreadas señalan los grupos entre los que tiene lugar la diferencia de promedios dentro del intervalo de confianza establecido ($p < .05$).

Es posible que estas diferencias guarden relación con la formación previa del profesorado (objetivo E2), también sujeta a varias conclusiones generalizables. Aunque en este caso los resultados obtenidos no se ciñen a un patrón fijo, en las diferencias apuntadas se advierte la especial presencia de los módulos de formación básica, por lo general impartidos por profesorado con formación previa en Pedagogía, Psicología y Psicopedagogía.

Las diferencias en la importancia atribuida al factor quinto y en la realización estimada para el tercero se producen entre los dos primeros módulos. Las/os estudiantes de módulos o asignaturas didáctico-disciplinares valoran la relevancia de la Adecuación y precisión evaluativa (Factor 5) en mayor medida que el alumnado que ha efectuado sus valoraciones en el módulo de formación básica ($p = .009$). En el cumplimiento percibido para el factor tercero (Ajustes en la evaluación) se produce a la inversa ($p = .017$).

Los promedios de cumplimiento total y los estimados para la Planificación de la evaluación (Factor 2) difieren dentro del intervalo de confianza establecido (IC = 95%) entre estudiantes del módulo didáctico disciplinar y los de mención y formación básica. En ambos casos, las respuestas inferiores las ofrece el alumnado que ha efectuado las valoraciones para módulos o asignaturas disciplinares.

Las valoraciones del alumnado de DECA sobre su Participación en el proceso evaluativo (Factor 1) son significativamente menores que las que proceden de estudiantes de otros módulos ($p \leq .047$). Las valoraciones registradas en los módulos de formación básica, didáctico disciplinar o de mención oscilan en torno al 2 (*poco adecuado*), mientras que las realizadas por el alumnado de DECA se sitúan por debajo de dicho nivel ($M = 1,6$).

La relevancia atribuida a la Regulación del aprendizaje por medio de la evaluación (Factor 4) difiere significativamente entre el alumnado que ha efectuado sus valoraciones en el módulo de mención y el que lo ha hecho en el de formación básica ($p = .007$). Los promedios resumidos en la tabla 133 indican que las valoraciones superiores las produce el primer grupo. Más complejas resultan las diferencias detectadas en la realización percibida,

que afectan a los cuatro grupos de diferente manera. En primer lugar, se evidencia que el alumnado que ha valorado sistemas evaluativos del módulo didáctico disciplinar considera menos realizado que el resto este cuarto factor ($p \leq .003$). Al mismo tiempo se advierte que las valoraciones producidas en el módulo DECA denotan una mayor regulación del aprendizaje que la que declaran estudiantes del de formación básica ($p = .003$).

Al puntuar el cumplimiento percibido en la Adecuación y precisión de la evaluación (factor quinto), las diferencias se producen en dos sentidos. Por una parte, los promedios registrados por el alumnado para el módulo didáctico disciplinar son significativamente menores que los que se constatan en mención ($p = .042$) y en formación básica ($p = .009$). Por otro lado, el promedio de realización en este último grupo es significativamente superior del que se produce en DECA ($p = .022$).

El Control de las condiciones evaluativas (sexto factor) se percibe realizado en distinta medida en tres grupos, resultando la diferencia de promedios significativa en el nivel establecido ($p < .05$). Aunque su interpretación requiera una cierta cautela, los resultados indican que el alumnado de formación básica advierte un mayor cumplimiento en este factor que el de DECA y el del módulo didáctico-disciplinar.

3.2.3 Influencias en la valoración de la competencia evaluativa.

Este último bloque integra una selección de factores de naturaleza diversa que pueden condicionar la valoración que realiza el alumnado de la competencia evaluativa del profesorado universitario con el que trabaja. En suma, analiza las percepciones de relevancia y realización de las/os estudiantes en función de cinco variables: la experiencia previa con el profesorado, el nivel de competencia docente atribuido, la existencia o no de resultados parciales de aprendizaje, las expectativas académicas en el módulo o asignatura valorada y la modalidad de respuesta llevada a cabo en el procedimiento de recogida de datos. Ordenadas en dos grupos, los cuatro primeros factores hacen referencia a algunas de las variables identificadas en la investigación sobre fuentes de sesgo en evaluaciones efectuadas por estudiantes, mientras que la quinta se asocia al propio proceso de investigación.

3.2.3.1 Fuentes de influencia potencial en las valoraciones del alumnado.

Uno de los factores que puede condicionar las valoraciones discentes sobre la competencia evaluativa del profesorado es la acumulación o no de una **experiencia previa** de trabajo en cursos y/o módulos/asignaturas anteriores. En los centros que conforman la muestra es frecuente que el alumnado trabaje con las/os mismas/os profesoras/es a lo largo de los cuatro años de titulación. Esta circunstancia es más probable cuanto más avanzado sea el curso, de modo que en la muestra conformada sólo se produce en el 20,6% de los casos⁹⁵.

Atendiendo a la existencia o no de una experiencia previa con la/el docente al cargo del sistema evaluativo valorado, las respuestas de la muestra difieren significativamente en el cumplimiento percibido en el Factor 1 (Participación discente en la evaluación) ($p = .021$) (ver tabla 134). La diferencia de promedios es muy leve y de efecto muy pequeño ($\eta^2 = 0,0076$); como resume la tabla 135, se produce en favor del alumnado que cuenta con un recorrido previo con la/el profesor/a de referencia, que reconoce mayores oportunidades de intervenir en la evaluación de su aprendizaje.

Tabla 134
Valoración de la competencia evaluativa docente en la muestra discente en función de la EXPERIENCIA PREVIA

Estadísticos de contraste (U de Mann Whitney)					
Factores	U de Mann-Whitney	W de Wilcoxon	Z	Sig. asintót. (bilateral)	TE (η^2)
Total IMPORTANCIA	36753,000	192714,000	-1,700	,089	
Total REALIZACIÓN	36918,500	192879,500	-1,623	,105	
FACTOR 1 IMP.	39046,000	195007,000	-,652	,514	
FACTOR 1 REA.	35441,000	191402,000	-2,311	,021	0,0076 MUY PEQUEÑO
FACTOR 2 IMP.	36829,500	192790,500	-1,682	,093	
FACTOR 2 REA.	39668,000	50253,000	-,362	,717	
FACTOR 3 IMP.	37544,000	193505,000	-1,373	,170	
FACTOR 3 REA.	38320,000	193723,000	-,579	,563	
FACTOR 4 IMP.	37822,000	193225,000	-1,223	,221	
FACTOR 4 REA.	32373,000	158626,000	-1,184	,236	
FACTOR 5 IMP.	36766,500	192169,500	-1,757	,079	
FACTOR 5 REA.	39595,000	50180,000	-,300	,765	

⁹⁵ Por razones de accesibilidad, el alumnado de primer curso representa casi el 60% de la muestra.

Estadísticos de contraste (U de Mann Whitney)

Factores	U de Mann-Whitney	W de Wilcoxon	Z	Sig. asintót. (bilateral)	TE (η^2)
FACTOR 6 IMP.	37612,000	193573,000	-1,328	,184	
FACTOR 6 REA.	37041,500	193002,500	-1,454	,146	

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. Los resultados ligados a este último factor se representan con letra cursiva para indicar la lectura cautelosa que requiere su interpretación por los indicios de ambigüedad en su estructuración interna.

Tabla 135

Diferencias significativas en las valoraciones de la muestra discente en función de la EXPERIENCIA PREVIA: estadísticos de grupo

Factores	Experiencia previa	n	M	s	SEM
F1. Participación REA.	Con experiencia	145	2,3603	,87890	,07299
	Sin experiencia	558	2,1876	,97256	,04117

La disposición o no de **resultados parciales** sobre su aprendizaje en el momento de llevar a cabo la valoración de la competencia evaluativa del profesorado puede constituir un segundo factor con influencia en la respuesta discente. Así lo confirma el contraste de promedios efectuado por medio de la U de Mann Whitney, que releva diferencias estadísticamente significativas ($p \leq .002$) entre ambos grupos de estudiantes, las/os que reconocen haber recibido algún resultado parcial de su trabajo en el módulo/asignatura valorada (el 66,3%) y aquellas/os otras/os que declaran no disponer de ninguna información en este sentido. Como refleja la tabla 136, todas ellas se producen en la subescala de realización con una incidencia de magnitud pequeña ($0,01 \leq \eta^2 \leq 0,06$):

Tabla 136

Valoración de la competencia evaluativa docente en la muestra discente en función de los RESULTADOS PARCIALES

Estadísticos de contraste (U de Mann Whitney)

Factores	U de Mann-Whitney	W de Wilcoxon	Z	Sig. asintót. (bilateral)	TE (η^2)
Total IMPORTANCIA	53041,500	80536,500	-,358	,720	
Total REALIZACIÓN	43088,000	70583,000	-4,338	,000	0,0271 PEQUEÑO
FACTOR 1 IMP.	52955,500	80450,500	-,396	,692	
FACTOR 1 REA.	50054,500	77549,500	-1,559	,119	
FACTOR 2 IMP.	51905,000	79400,000	-,821	,412	
FACTOR 2 REA.	39331,500	66826,500	-5,858	,000	0,0494 PEQUEÑO

Estadísticos de contraste (*U* de Mann Whitney)

Factores	<i>U</i> de Mann-Whitney	<i>W</i> de Wilcoxon	<i>Z</i>	Sig. asintót. (bilateral)	TE (η^2)
FACTOR 3 IMP.	53587,000	160078,000	-,144	,886	
FACTOR 3 REA.	50074,500	76639,500	-1,205	,228	
FACTOR 4 IMP.	51297,500	157788,500	-1,004	,316	
FACTOR 4 REA.	27131,000	42009,000	-6,119	,000	0,0593 PEQUEÑO
FACTOR 5 IMP.	52773,500	80034,500	-,395	,693	
FACTOR 5 REA.	45474,000	72270,000	-3,168	,002	0,0145 PEQUEÑO
<i>FACTOR 6 IMP.</i>	<i>52741,000</i>	<i>80236,000</i>	<i>-,487</i>	<i>,627</i>	
<i>FACTOR 6 REA.</i>	<i>52437,000</i>	<i>79698,000</i>	<i>-,512</i>	<i>,609</i>	

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. Los resultados ligados a este último factor se representan con letra cursiva para indicar la lectura cautelosa que requiere su interpretación por los indicios de ambigüedad en su estructuración interna.

En las cuatro valoraciones destacadas los promedios de cumplimiento superiores los concede el alumnado que ya dispone de resultados sobre su rendimiento (ver tabla 137). Debido a su propia naturaleza, esta tendencia resulta especialmente evidente en el factor cuarto, centrado en la Regulación del aprendizaje por medio de la evaluación. Además de concentrar la mayor discrepancia de promedios entre los dos grupos, la puntuación eta cuadrada registrada para el mismo también la destaca como la de mayor efecto ($\eta^2 = 0,0593$) (tabla 136).

Tabla 137

Diferencias significativas en las valoraciones de la muestra discente en función de los RESULTADOS PARCIALES: estadísticos de grupo

Factores	Resultados	<i>n</i>	<i>M</i>	<i>s</i>	<i>SEM</i>
Total REALIZACIÓN.	Con resultados	461	2,8979	,61259	,02853
	Sin resultados	234	2,6773	,61074	,03993
F2. Planificación REA.	Con resultados	461	3,1672	,74396	,03465
	Sin resultados	234	2,8021	,66862	,04371
F4. Regulación REA.	Con resultados	460	2,4529	,86014	,04010
	Sin resultados	172	1,9981	1,1031	,08412
F5. Adecuación... REA.	Con resultados	461	3,1236	,99777	,04647
	Sin resultados	231	2,8615	1,0541	,06936

Las **expectativas académicas del alumnado** en el módulo/asignatura considerada en la valoración discriminan tres subgrupos dentro de la muestra: el que prevé aprobar, el que cree que va a suspender y el que manifiesta dudas al respecto. Comparadas sus valoraciones por medio de la H de Kruskal Wallis, se aprecian seis diferencias significativas al nivel establecido, todas ellas de magnitud pequeña o muy pequeña (ver tabla 138):

Tabla 138

Valoración de la competencia evaluativa docente en la muestra discente en función de las EXPECTATIVAS ACADÉMICAS

Estadísticos de contraste (H Kruskal-Wallis)				
Factores	Chi-cuadrado	gl	Sig. asintót.	TE (η^2)
Total IMPORTANCIA	1,497	2	,473	
Total REALIZACIÓN	15,130	2	,001	0,0215 PEQUEÑO
FACTOR 1 IMP.	5,128	2	,077	
FACTOR 1 REA.	15,251	2	,000	0,0217 PEQUEÑO
FACTOR 2 IMP.	3,108	2	,211	
FACTOR 2 REA.	6,613	2	,037	0,0094 MUY PEQUEÑO
FACTOR 3 IMP.	,308	2	,857	
FACTOR 3 REA.	1,062	2	,588	
FACTOR 4 IMP.	7,498	2	,024	0,0107 PEQUEÑO
FACTOR 4 REA.	1,285	2	,526	
FACTOR 5 IMP.	5,291	2	,071	
FACTOR 5 REA.	16,926	2	,000	0,0241 PEQUEÑO
<i>FACTOR 6 IMP.</i>	<i>,195</i>	<i>2</i>	<i>,907</i>	
<i>FACTOR 6 REA.</i>	<i>11,892</i>	<i>2</i>	<i>,003</i>	<i>0,0169 PEQUEÑO</i>

Nota: El sombreado gris indica las valoraciones en las que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. Los resultados referidos a este último factor se representan con letra cursiva para indicar la cautela que requiere su interpretación.

Los análisis *post hoc* para el análisis de varianza permiten determinar entre qué grupos se producen las diferencias destacadas (ver tabla 139). Esta comprobación no ha sido posible para la realización asociada a la Planificación de la evaluación (Factor 2), dado que la ANOVA no capta en ella diferencias significativas ($p \geq .05$):

Tabla 139

Diferencias significativas en la valoración de la competencia evaluativa docente en función de las EXPECTATIVAS ACADÉMICAS del alumnado: estadísticos de grupo y contraste post hoc (T2 de Tamhane, T3 de Dunnet, C de Dunnet, Games-Howel, Bonferroni y Scheffé)

Diferencias significativas en función de las EXPECTATIVAS		<i>n</i>	<i>M</i>	<i>s</i>	Sig. Aprobar	Sig. Suspender	Sig. En duda
TOTAL REALIZACIÓN	Aprobar	528	2,8687	,60692	-	.565	.002
	Suspender	16	2,7022	,72515	.565	-	.982
	En duda	160	2,6715	,62491	.002	.982	-
FACTOR 1 REALIZACIÓN	Aprobar	528	2,3024	,96507	-	.381	.001
	Suspender	16	1,8750	1,1105	.381	-	.966
	En duda	160	1,9969	,86420	.001	.966	-
FACTOR 2 REALIZACIÓN	Aprobar	528	3,0726	,72318	-		
	Suspender	16	2,9500	,92232		-	
	En duda	160	2,9353	,75865			-
FACTOR 4 IMPORTANCIA	Aprobar	527	4,5117	,51037	-	.545	.044
	Suspender	16	4,1667	1,10219	.545	-	.321
	En duda	160	4,6229	,50034	.044	.321	-
FACTOR 5 REALIZACIÓN	Aprobar	526	3,1255	1,0101	-	.232	.000
	Suspender	16	2,6875	,87321	.232	-	.956
	En duda	159	2,7673	1,0184	.000	.956	-
FACTOR 6 REALIZACIÓN	Aprobar	527	3,4962	,83091	-	.855	.002
	Suspender	16	3,3333	,87771	.855	-	.947
	En duda	160	3,2177	,99369	.002	.947	-

Nota: Las celdas sombreadas señalan los grupos entre los que tiene lugar la diferencia de promedios dentro del intervalo de confianza establecido ($p < .05$).

Los resultados expuestos en la tabla 139 evidencian que las diferencias advertidas se producen entre el grupo de estudiantes que prevé aprobar y aquel otro que manifiesta dudas al respecto. El alumnado que anticipa un suspenso en el módulo/asignatura considerada (el 2,6%, el menos numeroso de la muestra) no difiere en sus valoraciones dentro del intervalo de confianza establecido.

En las valoraciones que hacen referencia al grado de realización los promedios superiores se concentran en el grupo de estudiantes con expectativas positivas. En la subescala de importancia la única diferencia con significación estadística ($p = .044$) se vincula a la Regulación del aprendizaje por medio de la evaluación (Factor 4) y se manifiesta de forma inversa. Es decir, en este caso quien atribuye mayor nivel de relevancia al factor es aquella parte del alumnado que alberga dudas sobre su calificación en el módulo/asignatura.

La **competencia docente percibida** se plantea como una variable *politómica*, en la que el alumnado califica al profesor/a de referencia en la valoración como poco competente, competente o muy competente. Los análisis efectuados a partir de la misma sitúan esta calificación como un factor relevante, que enlaza con una amplia gama de conclusiones generalizables a la población de referencia (IC = 95%). La Rho de Spearman evidencia que mantiene una correlación positiva y muy significativa ($p < .01$) con todas las valoraciones de la subescala de realización (ver tabla 140). De magnitud moderada (F2 y F5) o pequeña (Total y F1, F3, F4 y F6) esta asociación está sujeta a dos posibles interpretaciones, igualmente relevantes para la investigación en curso: cuanto mayor es la competencia, mayor cumplimiento se percibe en los factores analizados o a la inversa, cuanto más se llevan a cabo dichos factores, mayor es la atribución de competencia docente que efectúa el alumnado. Así sucede también para las puntuaciones de relevancia asignadas a la Planificación de la evaluación (F2), donde se detecta una correlación positiva, significativa ($p < .05$) y alta ($r_s = .079$).

Tabla 140

Correlación de la edad del alumnado y la valoración de la competencia evaluativa docente

COMPETENCIA DOCENTE	RHO de SPEARMAN		
	Coeficiente correlación	Sig. (bilateral)	N
TOTAL IMPORTANCIA	,014	,720	704
TOTAL REALIZACIÓN	,564	,000	704
F1. Participación en la evaluación IMP.	-,039	,298	704
F1. Participación en la evaluación REA.	,376	,000	704
F2. Planificación de la evaluación IMP.	,079	,036	704
F2. Planificación de la evaluación REA.	,530	,000	704
F3. Ajustes en la evaluación IMP.	,031	,405	704
F3. Ajustes en la evaluación REA.	,263	,000	700
F4. Regulación del aprendizaje IMP.	-,070	,063	703
F4. Regulación del aprendizaje REA.	,233	,000	641
F5. Adecuación y precisión IMP.	,015	,693	703
F5. Adecuación y precisión REA.	,442	,000	701
<i>F6. Control condiciones de evaluación IMP.</i>	,047	,212	704
<i>F6. Control condiciones de evaluación REA.</i>	,398	,000	703

Nota: El sombreado gris indica las valoraciones en las que se producen correlaciones significativas ($p < .05$) y el azul aquellas otras en las que la asociación detectada resulta muy significativa ($p < .01$). Los resultados referidos al factor sexto se representan con letra cursiva para indicar la cautela que requiere su interpretación.

La comparación de las respuestas producidas en los tres subgrupos de la muestra discente por medio de la H de Kruskal Wallis conduce a conclusiones coherentes con este marco de correlaciones (ver tabla 141). Nuevamente se constata que la realización percibida en los diversos factores varía significativamente en función del nivel de competencia docente atribuido al profesorado ($p = .000$). Salvo en el cumplimiento percibido en los factores tercero y cuarto, las puntuaciones eta cuadrada indican efectos de tamaño grande para estas diferencias ($\eta^2 \geq 0,14$), reforzando su relevancia.

Tabla 141

Valoración de la competencia evaluativa docente en la muestra discente en función de la COMPETENCIA DOCENTE ATRIBUIDA

Estadísticos de contraste (H Kruskal-Wallis)				
Factores	Chi-cuadrado	gl	Sig. asintót.	TE (η^2)
Total IMPORTANCIA	1,032	2	,597	
Total REALIZACIÓN	223,597	2	,000	0,3181 GRANDE
FACTOR 1 IMP.	2,986	2	,225	
FACTOR 1 REA.	99,558	2	,000	0,1416 GRANDE
FACTOR 2 IMP.	5,177	2	,075	
FACTOR 2 REA.	199,661	2	,000	0,2840 GRANDE
FACTOR 3 IMP.	2,173	2	,337	
FACTOR 3 REA.	48,719	2	,000	0,0697 MEDIO
FACTOR 4 IMP.	3,986	2	,136	
FACTOR 4 REA.	34,885	2	,000	0,0545 PEQUEÑO
FACTOR 5 IMP.	,627	2	,731	
FACTOR 5 REA.	138,188	2	,000	0,1974 GRANDE
<i>FACTOR 6 IMP.</i>	<i>1,857</i>	<i>2</i>	<i>,395</i>	
<i>FACTOR 6 REA.</i>	<i>111,895</i>	<i>2</i>	<i>,000</i>	<i>0,1594 GRANDE</i>

Nota: El sombreado gris indica las valoraciones en las que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. Los resultados referidos a este último factor se representan con letra cursiva para indicar la cautela que requiere su interpretación.

Replicado el análisis de los promedios por medio de ANOVA, las exploraciones *post hoc* determinan que las diferencias tienen lugar entre los tres grupos y siempre en la misma dirección (ver tabla 142). En el total de la subescala y en la realización atribuida a los seis factores, la puntuación más alta se sitúa siempre en el grupo que mayor competencia atribuye al docente responsable de la evaluación (muy competente) y la más baja en el grupo que lo define como poco competente.

Tabla 142

Diferencias significativas en la valoración de la competencia evaluativa docente en función de la COMPETENCIA DOCENTE ATRIBUIDA por el alumnado: estadísticos de grupo y contraste post hoc (T2 de Tamhane, T3 de Dunnet, C de Dunnet, Games-Howel, Bonferroni y Scheffé)

Diferencias sig. en función de la COMPETENCIA DOCENTE		<i>n</i>	<i>M</i>	<i>s</i>	Sig. Muy compt.	Sig. Competente	Sig. Poco compt.
TOTAL REALIZACIÓN	<i>Poco compet.</i>	90	2,2250	,37155	-	.000	.000
	<i>Competente</i>	403	2,7043	,53244	.000	-	.000
	<i>Muy compet.</i>	211	3,2951	,52537	.000	.000	-
FACTOR 1 REALIZACIÓN	<i>Poco compet.</i>	90	1,5861	,62840	-	.000	.000
	<i>Competente</i>	403	2,1071	,88483	.000	-	.000
	<i>Muy compet.</i>	211	2,7168	,97166	.000	.000	-
FACTOR 2 REALIZACIÓN	<i>Poco compet.</i>	90	2,4733	,43390	-	.000	.000
	<i>Competente</i>	403	2,8659	,64147	.000	-	.000
	<i>Muy compet.</i>	211	3,6097	,65666	.000	.000	-
FACTOR 3 REALIZACIÓN	<i>Poco compet.</i>	90	3,3370	1,2090	-	.026	.000
	<i>Competente</i>	399	3,7022	,99555	.026	-	.000
	<i>Muy compet.</i>	211	4,1785	,85943	.000	.000	-
FACTOR 4 REALIZACIÓN	<i>Poco compet.</i>	79	1,9156	,55333	-	.000	.000
	<i>Competente</i>	361	2,2359	,90938	.000	-	.000
	<i>Muy compet.</i>	201	2,6459	1,0430	.000	.000	-
FACTOR 5 REALIZACIÓN	<i>Poco compet.</i>	90	2,1722	,90926	-	.000	.000
	<i>Competente</i>	400	2,9225	,95066	.000	-	.000
	<i>Muy compet.</i>	211	3,6137	,85013	.000	.000	-
FACTOR 6 REALIZACIÓN	<i>Poco compet.</i>	90	2,7333	,93776	-	.000	.000
	<i>Competente</i>	402	3,3454	,80792	.000	-	.000
	<i>Muy compet.</i>	211	3,8855	,72715	.000	.000	-

Nota: Las celdas sombreadas señalan los grupos entre los que tiene lugar la diferencia de promedios dentro del intervalo de confianza establecido ($p < .05$).

Estos resultados son consistentes con la asociación hallada y refuerzan su significación desde la perspectiva teórica y empírica. Las respuestas obtenidas en la muestra discente sugieren que el profesorado al que se reconoce distinto nivel de competencia docente realiza la evaluación del aprendizaje de forma diferente. Las limitaciones del diseño de la investigación impiden determinar la direccionalidad a la que se ajusta dicha relación, pero su detección refuerza la necesidad y pertinencia de avanzar en la mejora de la praxis evaluativa en cualquier caso.

3.2.3.2 *El impacto de la modalidad de respuesta en los resultados hallados.*

Como ya ha sido justificado en el capítulo anterior, la modalidad de respuesta constituye una variable dicotómica creada a raíz de un imprevisto asociado a la recogida de datos en la muestra discente. El 85,2% de las/os participantes ha cumplimentado el cuestionario en una sesión dirigida de 45 minutos, contando con orientación expresa para minimizar la disparidad de interpretación en los indicadores valorados; el 14,8% restante ha respondido al cuestionario en una sesión de menor duración (30 minutos), lo que ha supuesto que la parte final del cuestionario se responda de manera autónoma (sin aclaración previa).

La comparación de las respuestas producidas en ambas modalidades de recogida (completa o abreviada) revela seis diferencias significativas dentro del intervalo de confianza establecido (IC = 95%) (ver tabla 143). Además de en el total de la subescala de realización, éstas tienen lugar en los factores primero, cuarto, quinto y sexto, precisamente aquellos cuyos indicadores se exponen en la parte final del cuestionario (la respondida de forma autónoma en la modalidad abreviada).

Tabla 143
Valoración de la competencia evaluativa docente en la muestra discente en función de la MODALIDAD DE RESPUESTA

Estadísticos de contraste (*U* de Mann Whitney)

Factores	<i>U</i> de Mann-Whitney	<i>W</i> de Wilcoxon	<i>Z</i>	Sig. asintót. (bilateral)	TE (η^2)
Total IMPORTANCIA	31049,000	36509,000	-,079	,937	
Total REALIZACIÓN	22087,500	202387,500	-4,760	,000	0,0322 PEQUEÑO
FACTOR 1 IMP.	30649,500	210949,500	-,290	,772	
FACTOR 1 REA.	20774,000	201074,000	-5,469	,000	0,0245 PEQUEÑO
FACTOR 2 IMP.	30561,000	210861,000	-,337	,736	
FACTOR 2 REA.	29461,500	209761,500	-,911	,362	
FACTOR 3 IMP.	29622,000	35082,000	-,847	,397	
FACTOR 3 REA.	28845,000	34305,000	-1,143	,253	
FACTOR 4 IMP.	30634,500	36094,500	-,279	,780	
FACTOR 4 REA.	22940,000	167393,000	-2,900	,004	0,0131 PEQUEÑO
FACTOR 5 IMP.	25196,500	204896,500	-3,291	,001	0,0154 PEQUEÑO
FACTOR 5 REA.	19874,500	198377,500	-5,924	,000	0,0501 PEQUEÑO
FACTOR 6 IMP.	30603,000	36063,000	-,317	,751	
FACTOR 6 REA.	24588,500	204288,500	-3,452	,001	0,0170 PEQUEÑO

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Los factores (F) se presentan en el mismo orden que han

sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. Los resultados ligados a este último factor se representan con letra cursiva para indicar la lectura cautelosa que requiere su interpretación por los indicios de ambigüedad en su estructuración interna.

La tendencia de respuesta en las diferencias que afectan a la subescala de realización se ajusta a un patrón claro con una única excepción. Salvo al valorar el grado de cumplimiento de la Participación discente en la evaluación (Factor 1), se advierte que el alumnado que ha respondido a una parte del cuestionario de manera autónoma concede puntuaciones significativamente superiores ($p \leq .004$) que quienes lo han hecho en la modalidad íntegramente guiada (ver tabla 144). Así se evidencia en el contraste de promedios para el Factor 6 (Control de las condiciones de la evaluación), el Factor 5 (Adecuación y precisión), el Factor 4 (Regulación del aprendizaje) o en el total de la subescala. En cambio, en la única diferencia significativa que afecta a la subescala de relevancia esta tendencia se invierte; al valorar la importancia del factor quinto es el grupo de estudiantes que ha participado de la modalidad completa quien ofrece valoraciones significativamente superiores ($p = .001$).

Tabla 144

Diferencias significativas en las valoraciones de la muestra discente en función de la MODALIDAD DE RESPUESTA: estadísticos de grupo

Factores	Modalidad	n	M	s	SEM
Total REALIZACIÓN.	Completa	600	2,7716	,60680	,02477
	Abreviada	104	3,1000	,61522	,06033
F1. Participación REA.	Completa	600	2,1328	,90725	,03704
	Abreviada	104	2,7452	1,0577	,10372
F4. Regulación REA.	Completa	599	4,5331	,52401	,02141
	Abreviada	104	4,5064	,57825	,05670
F5. Adecuación... IMP.	Completa	537	2,2719	,91483	,03948
	Abreviada	104	2,5994	1,0707	,10499
F5. Adecuación... REA.	Completa	597	2,9397	1,0053	,04114
	Abreviada	104	3,5769	,93676	,09186
<i>F6. Control evaluación REA.</i>	Completa	599	3,3809	,87733	,03585
	Abreviada	104	3,7067	,83430	,08181

Los resultados detectados en esta última sección representan un recordatorio de los sesgos que se producen en el proceso de recopilación de datos y evidencias. Aunque el tamaño de su efecto resulte pequeño ($\eta^2 \geq 0,06$), la modalidad de recogida produce

repercusiones estadísticamente significativas en las respuestas de la muestra, ya sea por contrarrestar la natural deseabilidad social o por incrementarla.

Breve avance de la discusión de resultados...

Los análisis efectuados en la resolución de este segundo objetivo han puesto de manifiesto un amplio abanico de relaciones e influencias entre factores socio-demográficos, profesionales y académico-institucionales vinculados a la muestra de estudiantes y la competencia evaluativa atribuida al profesorado. En términos generales las diferencias halladas resultan muy significativas (muy por debajo del intervalo de confianza establecido), aunque con un efecto limitado, principalmente de tamaño pequeño o muy pequeño. Las únicas variables que aparecen asociadas a efectos de magnitud grande son la materia y el nivel de competencia docente atribuido al profesor/a de referencia en la valoración.

Las limitaciones del diseño de la investigación no permiten arrojar conclusiones unilaterales, pero en las valoraciones discentes se evidencia una asociación importante entre la competencia docente atribuida al profesorado y el grado de cumplimiento percibido en los factores asociados a su competencia evaluativa. Esta conexión refuerza la incidencia de la evaluación del aprendizaje en la calidad formativa, ya sea como indicador o como requisito de la enseñanza competente. Así se constata también al valorar la calidad docente en las universidades públicas catalanas, donde la evaluación representa el tema que mayor número de comentarios genera por parte del alumnado (Rué, 2009b). Esta visión reafirma la incidencia y relevancia que se detalla para la evaluación desde la literatura (Biggs, 2005; Dochy et al., 2002; Gibbs & Dunbar-Goddet, 2007) y ofrece un testimonio consistente en favor de quienes advierten en ella un factor estratégico de calidad institucional (Gibbs, 2010; Kaneko, 2008; Porto, 2010; Villa, 2004).

El alumnado de la muestra reconoce distinto nivel de competencia evaluativa en el profesorado vinculado a las diferentes materias dentro del plan de estudios. Estas diferencias podrían estar relacionadas con las advertidas en la muestra docente en función de la formación previa (objetivo E2), y así lo sugieren los resultados detectados en otras dos variables próximas a ella (curso y módulo). Las valoraciones del alumnado refieren un mejor desempeño evaluativo en el módulo de formación básica y en el primer curso, donde mayor concentración se produce de tituladas/os en Pedagogía, Psicología y Psicopedagogía. Sin embargo, en torno a esta conclusión la investigación no revela resultados consistentes; indagando en el punto de vista discente, algunos estudios sugieren prácticas evaluativas diferentes en función del centro o área de conocimiento (V. Álvarez et al., 2002; Jessop & Maleckar, 2014; N. Martínez et al., 2012; Tejedor, 1999), mientras que en otros se descartan (Porto, 2006).

3.3 Resultados Objetivo F3. Identificar las principales fortalezas y debilidades en la evaluación del aprendizaje desde el punto de vista discente.

Este tercer sub-propósito completa desde una perspectiva más cualitativa la visión del alumnado sobre la competencia evaluativa docente. Para ello, parte del análisis de las fortalezas y áreas de mejora señaladas por las/os estudiantes para la evaluación en una pregunta final abierta del cuestionario. Respondida por el 70,5% de la muestra, las aportaciones recabadas desprenden una visión crítica del fenómeno evaluativo como ilustra la figura 43.

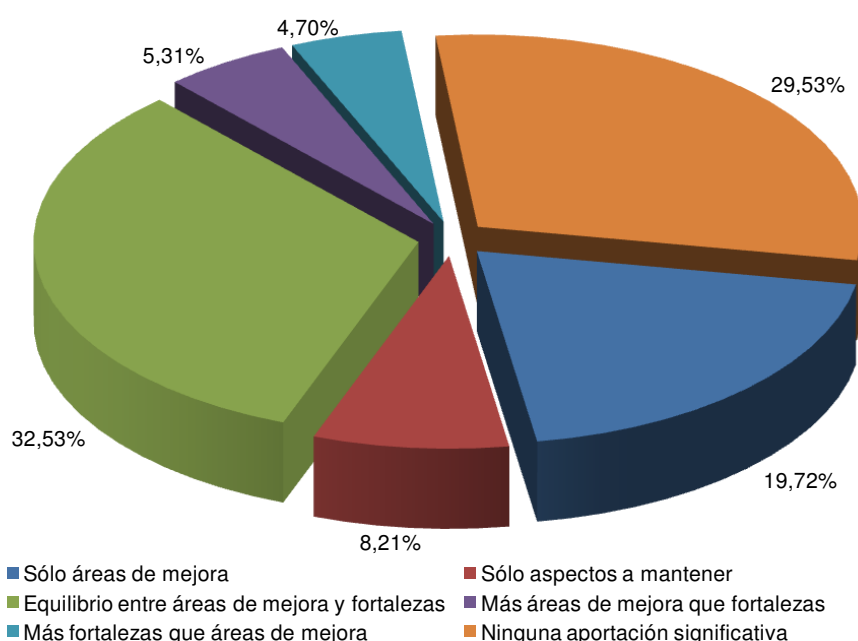


Figura 43. Resumen del tono general de las respuestas del alumnado en la pregunta referida a las fortalezas y las áreas de mejora en la evaluación de su aprendizaje.

El total de aportaciones efectuadas a esta cuestión asciende a 1128, 641 de ellas (el 56,82%) referidas a las áreas de mejora y a fortalezas las 487 restantes (43,17%). En ambos casos se evidencia que las contribuciones y comentarios del alumnado se encuentran estrechamente relacionados con los aspectos tratados en el cuestionario, lo que podría estar indicando una cierta transformación de la muestra. Según Manzano y Braña (2005) la participación en la investigación cambia a las personas, y así ha podido suceder a partir de la cumplimentación guiada del cuestionario. En consecuencia, sería interesante poder comprobar si pasado cierto tiempo, las/os estudiantes señalan las mismas deficiencias y bondades para la praxis evaluativa.

Las contribuciones del alumnado sobre las fortalezas de la evaluación han sido sintetizadas en las 11 categorías presentadas en la tabla 145. Ordenadas en función de la frecuencia, las tres primeras adoptan un evidente protagonismo en el banco de respuestas:

Tabla 145

Fortalezas en la evaluación del aprendizaje según el alumnado de la muestra

FORTALEZAS (aspectos a mantener)	f	%
Evaluación continua	161	33,05
Evaluación variada	139	28,54
Información clara sobre la evaluación del aprendizaje	93	19,09
Relación entre el trabajo cotidiano y las demandas de la evaluación	19	3,90
Evaluación igual para todas/os	17	3,49
Posibilidad de rehacer trabajos a partir de las correcciones	16	3,28
Realización de trabajos/proyectos grupales	13	2,66
Revisiones de trabajos y exámenes (a demanda, tutoría)	9	1,84
Consideración de la participación en el aula	9	1,84
Correcciones colectivas o demostraciones	7	1,43
Nivel de dificultad accesible (fácil aprobar)	4	0,82
TOTAL DE FORTALEZAS IDENTIFICADAS	487	

El 33,05% de las aportaciones recogidas señalan la *continuidad de la evaluación* como uno de los aspectos a mantener en la práctica. Vinculándola a la función acreditativa, las/os estudiantes indican que es positivo que su calificación se decida a partir de tareas o pruebas realizadas a lo largo de todo su proceso formativo, no sólo en la recta final.

Estrechamente relacionado con la continuidad, el alumnado refiere que la *variedad de formatos evaluativos* representa otra de las fortalezas de la evaluación desarrollada por el profesorado. En este caso, son frecuentes las alusiones a la combinación de las clásicas pruebas escritas (exámenes) con otro tipo de tareas evaluativas, fundamentalmente trabajos, proyectos o prácticas de aula.

El tercer tema más representado entre las aportaciones discentes es la *claridad de las condiciones de la evaluación* (19,09%). Las respuestas recopiladas describen los beneficios de contar con información sobre el sistema evaluativo, particularmente sobre el tipo de pruebas o tareas que determinan la calificación y sobre los criterios aplicados en su valoración. Aunque aparece referida de forma minoritaria (7 registros, el 1,43% del total), las *correcciones colectivas en el aula o las demostraciones* también guardan relación con esta claridad.

Los tres grandes núcleos de fortalezas evaluativas o aspectos a mantener se corresponden de forma directa con el contenido de tres de los ítems incluidos en el cuestionario (ver figura 44). Los promedios de importancia registrados para los mismos son similares; denotan que constituyen cuestionespreciadas para la muestra, pero no se sitúan en el mismo nivel de realización. La *variedad en la evaluación* (recogida en el ítem C.1.4. referido a los instrumentos), se considera adecuadamente cubierta en la práctica ($M = 3,37$), de modo que parece lógica su inclusión entre las fortalezas. La *continuidad en la evaluación* también se considera realizada a un nivel *adecuado* ($M = 2,94$, casi en la media teórica), pero en el caso de la información y la *claridad de las condiciones evaluativas* no sucede lo mismo. Explicado presumiblemente por el elevado nivel de dispersión que concentra ($s = 1,13^{96}$), los promedios registrados reflejan se trata de un indicador que no alcanza un cumplimiento *adecuado* en la práctica ($M < 3$), aunque una parte del alumnado lo señale como fortaleza.

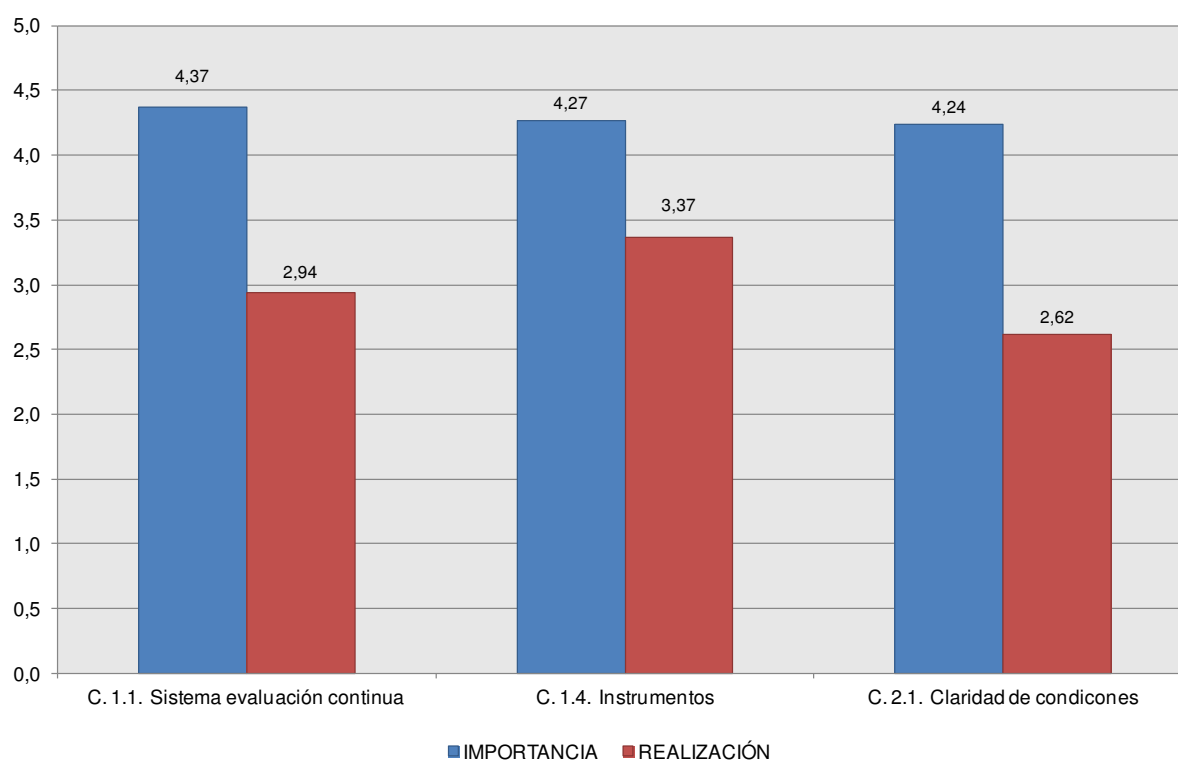


Figura 44. Representación de las tres principales fortalezas evaluativas señaladas por la muestra discente con los datos recabados por medio del cuestionario.

El resto de fortalezas obtienen un nivel de representación mucho menor en las respuestas del alumnado. El 3,9% de las aportaciones hacen referencia a la *relación de*

⁹⁶ La información completa se expone en el **ANEXO 28**.

alineación entre el trabajo cotidiano y las exigencias evaluativas. En coherencia, el ítem que recoge esta cuestión concreta (C.1.2.) destaca como el más valorado en importancia y se le asocia un nivel de cumplimiento por encima de la media teórica ($M = 3,48$, según detalla el **ANEXO 28**).

La necesidad y pertinencia de adecuar las condiciones de la evaluación en respuesta a circunstancias particulares no representa un acuerdo extendido entre el alumnado⁹⁷. Así lo evidencia nuevamente otra de las fortalezas recopiladas, que señala la *igualdad de las condiciones de la evaluación* como aspecto a mantener en la práctica (3,49% de las aportaciones).

Relacionada con la función reguladora de la evaluación, la *posibilidad de rehacer trabajos* a partir de las pautas y correcciones derivadas de la evaluación representa una fortaleza, aunque poco extendida (3,28% de las respuestas). De hecho, su presencia es mucho mayor en el próximo bloque, donde figura como una de las más reclamadas áreas de mejora.

Esta enumeración de las principales fortalezas percibidas por las/os estudiantes se completa con otras cuestiones menos presentes entre sus respuestas. Ente ellas se encuentran la *oportunidad de realizar trabajos o proyectos grupales* (2,66%) y la *revisión de trabajos y/o exámenes fuera del aula* (1,84%). La accesibilidad en el nivel de *exigencia de la asignatura/módulo* y la incidencia de la *asistencia y participación en clase* en la calificación también figuran en las prácticas destacadas (ocupan el 0,82% y el 1,84% de las mismas, respectivamente).

En cuanto a las áreas de mejora, el análisis y clasificación de las aportaciones realizadas por la muestra discente ha dado lugar a la categorización expuesta en la tabla 146. En suma se detectan 11 grandes ejes de mejora, con especial representación de los dos primeros, entre los que los que presumimos una estrecha relación:

Tabla 146
Áreas de mejora en la evaluación del aprendizaje según las respuestas abiertas del alumnado

ÁREAS DE MEJORA	f	%
Feedback deficiente	297	46,33
⇒ Impuntualidad	71	11,07
⇒ Insuficiente detalle e individualización	226	35,25

⁹⁷ Considerada inicialmente, la inconsistencia y dispersión de las respuestas del alumnado llevaron a suprimirla de la versión definitiva del instrumento, según se ha detallado en el capítulo anterior.

ÁREAS DE MEJORA	f	%
Ausencia de resultados parciales	167	26,05
Sobrecarga final	45	7,02
Agravios comparativos (injusticias, tratos de favor)	39	6,08
Falta de información sistema evaluativo	22	3,43
Imprecisión, arbitrariedad de criterios	20	3,12
Coordinación de agendas en exigencias evaluativas	14	2,18
Participación en metaevaluación	13	2,02
Evaluación no válida	11	1,71
Rigidez del sistema evaluativo	8	1,24
Evaluación de trabajos en equipo	5	0,78
TOTAL DE OBSTÁCULOS IDENTIFICADOS		938

Según las respuestas recogidas, el aspecto que mayor insatisfacción produce entre el alumnado es el *proceso de feedback o retroalimentación*, sobre el que se cuentan 297 comentarios. El 11,07% señala que los resultados llegan excesivamente tarde, limitando las oportunidades reales de mejora y la orientación sobre la propia evolución. En el 35,25% restante se describe una comunicación de resultados insuficiente, que no ayuda a comprender los errores y lagunas constatados, ni ayuda a identificar claves de mejora (información fundamentalmente cuantitativa de resultados y comentarios vagos y generales sobre los errores cometidos). Frente a ello, el alumnado demanda fórmulas de feedback individualizado que ofrezcan comentarios concretos sobre los aspectos a subsanar y la forma de mejorarlos.

Ligado al anterior, el segundo núcleo de mejora hace referencia a la *ausencia de resultados parciales*. El 26% de las aportaciones del alumnado señalan el desconocimiento o la desinformación de los resultados de aprendizaje y el avance efectuado en el proceso formativo, ya sea porque no se ha llevado a cabo ninguna prueba o tarea específica para su determinación o porque aún no se disponga de información sobre los resultados. Efectuada aproximadamente a dos semanas del fin del cuatrimestre, en la recopilación de datos de identificación efectuada el 33,7% de la muestra discente declara no contar con ninguna información en este sentido, lo que reafirmaría la debilidad detectada.

Con mucha menor representación que las dos anteriores, la tercera área de mejora propuesta tiene que ver con la gestión práctica del principio de la evaluación continua. Incluso figurando entre las fortalezas referidas, las/os estudiantes consideran que las tareas y pruebas evaluativas se distribuyan a lo largo del proceso formativo, aunque con una *concentración desproporcionada* de las mismas en el tramo final. El impacto de esta

deficiencia se incrementa para el alumnado ante la *falta de coordinación en las exigencias evaluativas del profesorado* de las diferentes asignaturas o módulos, circunstancia señalada en el 2,18% de las aportaciones.

El 6% de los comentarios recibidos presentan *las injusticias o los agravios comparativos* como aspecto a mejorar en la evaluación. En ellos se describen beneficios o ventajas selectivas para determinadas/os estudiantes en función de circunstancias personales o académicas (inferior competencia lingüística, cargas familiares o laborales, convocatorias agotadas, etc.). Esta limitación remite nuevamente al controvertido principio de la atención a la diversidad en la evaluación; en los comentarios registrados se produce una cierta asimilación entre ajuste y agravio, mientras que la igualdad de condiciones se señala como fortaleza.

Aunque la claridad de las condiciones evaluativas figura entre las principales fortalezas reconocidas en la evaluación, se producen 22 comentarios (3,43%) apuntando la necesidad de *mejorar la información sobre los sistemas de evaluación aplicados*. En similar proporción se hace referencia a la *arbitrariedad o falta de fiabilidad* de la evaluación, donde se describen correcciones poco rigurosas, influenciadas por las expectativas y prejuicios del profesorado y basadas en criterios poco claros.

Acaso por influencia de la propia cumplimentación del cuestionario, el 2% de las aportaciones señalan como área de mejora la participación de las/os estudiantes en la revisión de la evaluación, esto es, en la *metaevaluación*. Esta apreciación es coherente con las valoraciones que ha realizado la muestra para los ítems del cuestionario centrados en la metaevaluación (E.8.1., E.8.2. y E.8.3., referidos a la revisión pedagógica, técnica y ética, respectivamente).

Para completar el listado de áreas de mejora, queda reseñar la *falta de validez* de los sistemas evaluativos, que se sugiere en 11 ocasiones (el 1,71%). Las observaciones constatas en este sentido, señalan prácticas evaluativas que no muestran verdaderamente aprendizaje y que se centran prioritariamente en la reproducción de contenidos teóricos intrascendentes. Por otra parte, las *injusticias asociadas a la valoración de productos grupales* y la *rigidez de los dispositivos evaluativos* también encuentra un tímido reflejo entre las sugerencias de mejora producidas (el 0,78% y el 1,24% de las respuestas, respectivamente). Relacionado con la adecuación de las condiciones evaluativas, este segundo aspecto reclama flexibilidad para la evaluación (asistencia irregular a clase, pruebas alternativas, etc.), aunque sólo de forma minoritaria.

Breve avance de la discusión de resultados...

Las fortalezas y debilidades que el alumnado aprecia en la evaluación del aprendizaje guardan una estrecha conexión con las valoraciones realizadas y con los indicadores tratados en el análisis. Esa coincidencia puede ser interpretada como signo de reafirmación en las mismas o como indicio de cierto sesgo por transformación de la muestra (Manzano & Braña, 2005).

Las/os estudiantes valoran los aspectos ligados a la planificación evaluativa como los más llevados a cabo y como fortalezas que conviene mantener. En contraste con la imagen que devolvían diagnósticos efectuados a lo largo de la década de los 90 (M. Fernández, 1989; Flórez, 1999; Tejedor, 1999), el alumnado valora positivamente la continuidad de la evaluación, la variedad de instrumentos y la claridad de información recibida sobre los criterios y condiciones. Así lo constatan también trabajos basados en el análisis de los programas docentes y la normativa evaluativa (Ibarra & Rodríguez, 2010; Padilla et al., 2010; M. A. Sánchez et al., 2010) o en el punto de vista del profesorado y las/os gestoras/es de formación (Villa et al., 2013). En aquellos otros trabajos que consultan directamente la perspectiva discente se advierten conclusiones diferentes dependiendo del aspecto concreto de la planificación al que hagan referencia. La continuidad aparece entre los rasgos más ampliamente reconocidos en la evaluación, aunque desde una visión prioritariamente acreditativa (Hamodi & López, 212; V. M. López, 2008). Esta misma tendencia se observa al valorar la variedad de herramientas aplicables, ya que se admite la diversidad metodológica, pero no su aplicación normalizada (Gutiérrez et al., 2011; L. F. Martínez et al., 2012; Martínez-Mínguez & Flores, 2014). La valoración de la claridad y transparencia evaluativa representa uno de los aspectos que mayor controversia suscita entre el alumnado, según los resultados reportados en la investigación. La información sobre el proceso de evaluación destaca como uno de los aspectos mejor valorados en unos casos (Buckley, 2012; Gil-Flores, 2012; Rué, 2009b), representando en otros motivo de insatisfacción y queja (Arribas et al., 2010; Burgess, 2004; V. M. López & Palacios, 2012).

Confirmando las conclusiones de otros estudios sobre la praxis evaluativa en la universidad, el feedback ocupa un lugar destacado entre las áreas de mejora referidas (Buckley, 2012; Ferguson, 2011; Gutiérrez et al., 2013; G. Rodríguez et al., 2012). En coherencia con las orientaciones teóricas definidas en el modelo ITC, el alumnado demanda más feedback, más puntual y más personalizado. Las sugerencias de cambio recogidas insisten en la necesidad de dotar la retroalimentación de carácter prospectivo, y así se reivindica también en la literatura especializada (Hounsell, 2003; Nicol, 2013; Sadler, 2010; Orsmond et al., 2013; Walker, 2013). Esta necesidad es compartida por el profesorado, que reclama condiciones prácticas que le permitan una mejor atención a este espacio de la evaluación que considera prioritario (objetivo F3 y F1).

4. Resultados relativos al Objetivo G. Contraste de perspectivas sobre la competencia evaluativa docente

Analizada en los dos objetivos previos la visión que sostienen de la competencia evaluativa estudiantes y profesorado, este nuevo apartado plantea el contraste de las valoraciones constatadas en ambas secciones de la muestra. Dicho contraste sólo se ha efectuado sobre los aspectos comparables, ya que cada fuente ha cumplimentado un cuestionario específico, dando lugar a una estructuración desigual en sus respuestas (**ANEXO 27**):

- Como ya ha sido detallado en diversas ocasiones, los ítems comunes en ambas escalas son aquellos que incluyen una C (común) en su codificación. Eso reduce a 9 la lista de indicadores directamente comparables (C.1.1., C.1.2., C.1.3., C.1.4., C.2.1., C.5.3., C.6.1., C.6.2., C.7.1.). A pesar de ser comunes, quedan fuera de la comparativa los indicadores constituyentes de la dimensión Atención a la diversidad en el modelo teórico inicial (C.3.1. y C.3.2.) por haber sido suprimidos del cuestionario discente para mejorar su estructura interna.
- En cuanto a la estructuración de estos nueve elementos, el contraste directo de los puntos de vista docente y discente se realiza principalmente de acuerdo a la división original prevista en el modelo teórico sobre la competencia evaluativa:
 - ✓ **Dimensión primera del modelo teórico (D1)**, relativa a la Planificación estratégica de la evaluación del aprendizaje. Dicha dimensión se corresponde con el Factor 1 del cuestionario docente que lleva el mismo nombre y con el Factor 2 del discente (Planificación de la evaluación) si se le suprime el ítem C.2.1., el que en menor medida correlaciona con el resto. Los elementos que integran esta primera dimensión comparable son cuatro (C.1.1., C.1.2., C.1.3. y C.1.4.).
 - ✓ **Dimensión segunda del modelo teórico (D2)** sobre la Claridad de las condiciones y criterios de la evaluación. Dicha dimensión, se corresponde con el ítem C.2.1., incluido dentro de otros factores en ambas secciones de la muestra.
 - ✓ **Dimensión séptima del modelo teórico (D7)**, relativa a la Participación de las/os estudiantes en la evaluación del aprendizaje. Dicha dimensión es la segunda conformada por un único elemento (C.7.1.).

✓ **Indicadores C.5.3., C.6.1. y C.6.2. (Dimensión quinta redefinida y común)**

(D5-II). Los tres indicadores restantes se corresponden con la nueva delimitación de la dimensión quinta en la versión revisada de la rúbrica de referencia (**ANEXO 20**). Ésta coincide con el Factor 4 del cuestionario discente (Regulación del aprendizaje por medio de la evaluación) y con el Factor 2 del aplicado al profesorado (Regulación del aprendizaje y la enseñanza por medio de la evaluación). Para su comparación directa en ambos sectores de la muestra, en este caso no se considera el indicador 5.4 (ítem P.6.3. en la versión original), que es exclusivo para docentes al referirse a la regulación de la enseñanza.

- En el resto de dimensiones o factores las valoraciones de docentes y discentes sólo se pueden comparar indirectamente. Descartados los nueve indicadores comunes, los ítems restantes remiten a las dimensiones cuarta (Implementación del sistema evaluativo), quinta (Análisis e interpretación de evidencias) y octava (Metaevaluación) del modelo teórico inicial. Los ítems integrados en las mismas están formulados desde puntos de vista diferentes, de modo que sólo se pueden interpretar en paralelo o a modo de contraste.

Un primer dato destacable para introducir el contraste de visiones es que ambos colectivos coinciden en situar en el área de Concentración de la representación AIR todas las dimensiones comparables de la competencia evaluativa (ver figura 45).

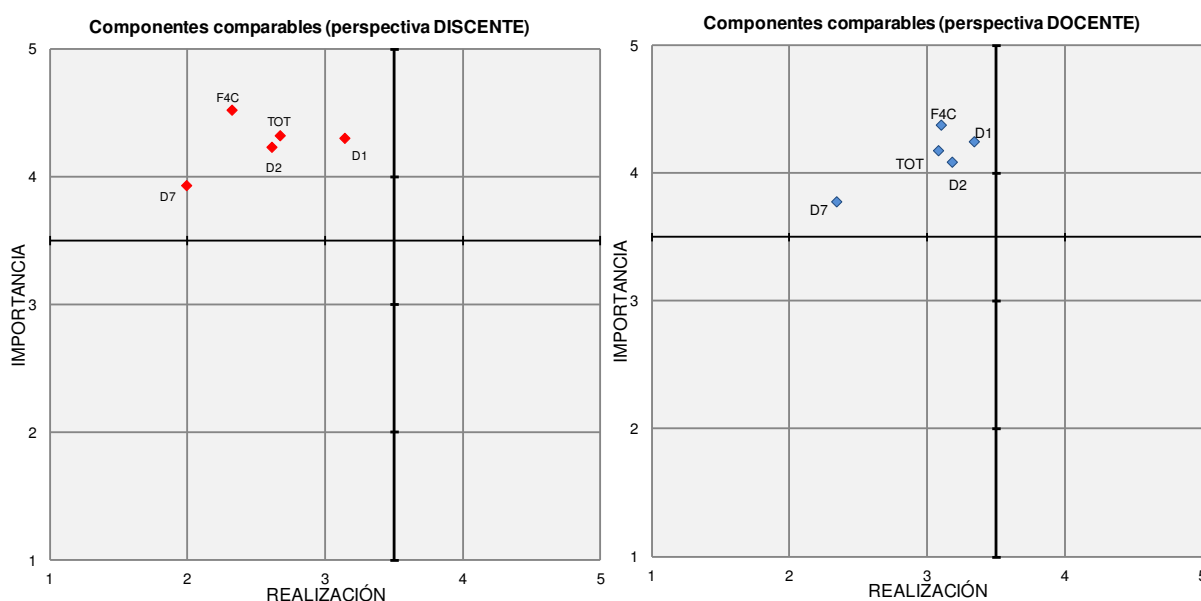


Figura 45. Perspectiva docente y discente en los componentes comparables en la valoración de la competencia evaluativa.

D1 (C.1.1., C.1.2., C.1.3., C.1.4.), D2 (C.2.1.), D7 (C.7.1.), D5-II (C.5.3., C.6.1., C.6.2.) y TOT. (Total de los nueve indicadores). Las valoraciones docentes están representadas en rojo y en el gráfico de la izquierda. La respuesta de la muestra de estudiantes se representa a la derecha y en azul.

Tanto para estudiantes como para profesorado, se produce un claro desajuste entre la relevancia de los componentes asociados a la competencia evaluativa y su grado de cumplimiento práctico, aunque no en igual medida. Así lo evidencia esta primera comparación de perspectivas, que anticipa la tendencia general de respuesta que se constata para cada una de las dimensiones contrastadas en próximas páginas. Frente a las puntuaciones explicitadas en la muestra docente, el alumnado otorga mayores valoraciones al calificar la relevancia y menores al juzgar la realización. Salvada dicha tendencia, las valoraciones realizadas en ambas fuentes presentan ciertas similitudes:

- Tanto estudiantes como docentes advierten que la Participación discente en la evaluación (D7) resulta la dimensión menos relevante y llevada a cabo.
- En ambas representaciones la dimensión que más se aproxima al área de Mantenimiento es la primera, aquella que refiere a la Planificación estratégica de la evaluación (D1).
- La Regulación del aprendizaje por medio de la evaluación (D5-II) representa para docentes y discentes la dimensión más relevante de la competencia evaluativa, por eso ocupa el punto de mayor altura en ambas representaciones.
- En ambos sectores de la muestra la Claridad de los criterios y condiciones de la evaluación (D2) representa el segundo aspecto de menor relevancia en ambas muestras.

Los próximos apartados profundizan en esta perspectiva general delimitada para las diferencias, indagando sobre su posible generalización a la población de referencia. El primero de ellos atiende a los resultados registrados en la subescala de importancia y el segundo a los constatados en la percepción de realización.

4.1 Valoraciones de importancia en los factores y dimensiones vinculadas a la competencia evaluativa docente.

Comparados por medio de la U de Mann Whitney (ver tabla 147), los promedios de relevancia registrados en ambas secciones de la muestra se constata una única diferencia significativa al nivel establecido ($p < .05$).

Tabla 147

Contraste de promedios de importancia de estudiantes y profesorado en las dimensiones comparables de la competencia evaluativa

Estadísticos de contraste (U de Mann Whitney)

IMPORTANCIA	U de Mann-Whitney	W de Wilcoxon	Z	Sig. asintót. (bilateral)	TE (η^2)
D1. Planificación	17499,500	18825,500	-,306	,760	
D2. Claridad	15953,000	17279,000	-1,449	,147	
D7. Participación	16597,000	17923,000	-,969	,332	
D5 (II). Regulación	14760,000	16086,000	-2,187	,029	0,0063 MUY PEQUEÑO
TOTAL	15204,500	16530,500	-1,833	,067	

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Las dimensiones comparadas son D1. (Planificación estratégica de la evaluación), D2 (Claridad de las condiciones y criterios de la evaluación), D7 (Participación discente en la evaluación), D5-II (Regulación del aprendizaje por medio de la evaluación redefinida y común) y TOTAL.

Al valorar la importancia de la Regulación del aprendizaje por medio de la evaluación (D5-II) el alumnado ofrece puntuaciones significativamente mayores que el profesorado ($p = .029$). En correspondencia con el tamaño del efecto estimado para la diferencia (TE muy pequeño), la distancia entre ambos promedios de respuesta resulta muy reducida (ver tabla 148), aunque generalizable a la población.

Tabla 148

Diferencias significativas en el contraste de los promedios de importancia comparables en función de la FUENTE: estadísticos de grupo

Factores	Fuente	n	M	s	SEM
D5 (II). Regulación	Docente	51	4,3856	,54721	,07662
	Discente	703	4,5292	,53203	,02007

Analizadas las valoraciones registradas para los ítems que integran dicha dimensión, las diferencias significativas se reducen a dos de ellos, ambas con un efecto de magnitud pequeña ($\eta^2 < 0,06$) según refleja la tabla 149:

Tabla 149

Contraste de promedios de importancia de estudiantes y profesorado en los componentes de la Regulación del aprendizaje por medio de la evaluación (D5 – II)

Estadísticos de contraste (U de Mann Whitney)

IMPORTANCIA	U de Mann-Whitney	W de Wilcoxon	Z	Sig. asintót. (bilateral)	TE (η^2)
C.5.3. Resultados claros, confidenciales y concisos	14117,500	15443,500	-2,923	,003	0,0114 PEQUEÑO
C. 6.1. Oferta feedback	17731,500	265187,50	-,149	,881	
C. 6.2. Pautas proacción	14364,000	15690,000	-2,750	,006	0,0100 PEQUEÑO

La importancia de la *comisión de feedback* a partir de las actividades evaluativas desarrolladas (ítem C.6.1.) no da lugar a valoraciones significativamente diferentes en función de la fuente ($p = .881$). Éstas sólo difieren dentro del intervalo de confianza establecido al valorar la *presentación clara, concisa y confidencial de resultados de aprendizaje* (C. 5.3.) ($p = .003$) y la proporción de *pautas de proacción* (C. 6.2.) ($p = .006$). Como recoge la tabla 150, en ambos casos, las estimaciones del alumnado se sitúan ligeramente por encima de las del profesorado.

Tabla 150

Diferencias significativas en los componentes de la Regulación del aprendizaje por medio de la evaluación (D5 – II) en función de la FUENTE: estadísticos de grupo

Factores	Fuente	n	M	s	SEM
C.5.3. Resultados claros, confidenciales y concisos	Docente	51	4,29	,701	,098
	Discente	701	4,54	,687	,026
C. 6.2. Pautas proacción	Docente	51	4,33	,622	,087
	Discente	703	4,55	,636	,024

4.2 Valoraciones de realización percibida en los factores y dimensiones vinculadas a la competencia evaluativa docente.

El contraste de los promedios de realización percibida en ambos sectores de la muestra da lugar a los resultados expuestos en la tabla 151. En este caso, las valoraciones de estudiantes y profesorado difieren dentro del intervalo de confianza establecido (IC = 95%) en todos los componentes de la competencia evaluativa comparados. Las puntuaciones eta cuadrada asociadas a tales diferencias estiman una magnitud de efecto pequeña o muy pequeña ($\eta^2 < 0,06$).

Tabla 151

Contraste de promedios de realización percibida en estudiantes y profesorado para las dimensiones comparables de la competencia evaluativa

Estadísticos de contraste (U de Mann Whitney)

REALIZACIÓN	U de Mann-Whitney	W de Wilcoxon	Z	Sig. asintót. (bilateral)	TE (η^2)
D1. Planificación	14401,000	262561,000	-2,160	,031	0,0062 MUY PEQUEÑO
D2. Claridad	12211,000	259667,000	-3,747	,000	0,0187 PEQUEÑO
D7. Participación	13826,500	257777,500	-2,591	,010	0,0090 MUY PEQUEÑO
D5 (II). Regulación	8149,500	213910,500	-5,825	,000	0,0492 PEQUEÑO
TOTAL	10958,500	259118,500	-4,466	,000	0,0265 PEQUEÑO

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Las dimensiones comparadas son D1. (Planificación estratégica de la evaluación), D2 (Claridad de las condiciones y criterios de la evaluación), D7 (Participación discente en la evaluación), D5-II (Regulación del aprendizaje por medio de la evaluación redefinida y común) y TOTAL.

En todas las diferencias detectadas los promedios se ajustan al patrón de respuesta anunciado en la introducción (ver tabla 152). Las mayores puntuaciones las otorga el profesorado, que en su propia praxis evaluativa aprecia un cumplimiento de estas dimensiones significativamente mayor que el que describe en sus valoraciones el alumnado ($p < .05$). La única dimensión que sitúan las/os estudiantes por encima de la media teórica de realización es la Planificación estratégica de la evaluación (D1); en evidente contraste, los promedios de valoración docente sólo se colocan por debajo de dicho límite al considerar el cumplimiento de la Participación discente en la evaluación (D7).

Las mayores distancias se producen en la valoración del grado de cumplimiento de la función reguladora de la evaluación (D5-II) y el grado de Claridad de los criterios y condiciones de la evaluación (D2) en la práctica. Menor es la disparidad de percepciones al valorar la realización de la Planificación estratégica de la evaluación (D1), donde con una

magnitud de efecto muy pequeña ($\eta^2 = 0,0062$), la diferencia de promedios apenas alcanza los 0,20 puntos entre ambas fuentes.

Tabla 152

Diferencias significativas en el contraste de los promedios de realización de las dimensiones comparables en función de la FUENTE: estadísticos de grupo

Factores	Fuente	n	M	s	SEM
D1. Planificación	Docente	50	3,3450	,71230	,10073
	Discente	704	3,1437	,74865	,02822
D2. Claridad	Docente	50	3,1800	,89648	,12678
	Discente	703	2,6174	1,1310	,04266
D7. Participación	Docente	50	2,3400	1,0022	,14174
	Discente	698	1,9928	1,0666	,04037
D5 (II). Regulación	Docente	50	3,1000	,65379	,09246
	Discente	641	2,3250	,94869	,03747
TOTAL	Docente	50	3,0800	,54173	,07661
	Discente	704	2,6729	,66586	,02510

La tabla 153 presenta los resultados del contraste de promedios de cumplimiento percibido en cada uno de los indicadores que conforman las dimensiones comparables. Este segundo análisis revela que la diferencia significativa de promedios señalada para la Planificación estratégica (D1) se limita a la percepción que manifiestan estudiantes y profesorado sobre la *continuidad de la evaluación* (ítem C.1.1.) ($p = .002$). No sucede así en la Regulación del aprendizaje por medio de la evaluación (D5-II), donde ambos sectores de la muestra declaran percepciones significativamente diferentes en el cumplimiento de los tres indicadores asociados, todos ellos con una magnitud pequeña en el efecto estimado ($p = .000$; $\eta^2 < 0,0384$).

Tabla 153

Contraste de promedios de realización percibida en estudiantes y profesorado para los indicadores comparables de la competencia evaluativa

Estadísticos de contraste (U de Mann Whitney)

REALIZACIÓN	U de Mann-Whitney	W de Wilcoxon	Z	Sig. asintót. (bilateral)	TE (η^2)
C.1.1. Sistema evaluación continua	13166,500	260622,50	-3,082	,002	0,0126 PEQUEÑO
C.1.2. Propuesta evaluativa alineada	15350,500	262103,50	-1,600	,110	
C.1.3. Combinación formativa y sumativa	15223,500	263383,50	-1,640	,101	

Estadísticos de contraste (*U* de Mann Whitney)

REALIZACIÓN	<i>U</i> de Mann-Whitney	<i>W</i> de Wilcoxon	<i>Z</i>	Sig. asintót. (bilateral)	TE (η^2)
C.1.4. Previsión instrumentos	15497,000	16772,000	-1,429	,153	
C.2.1. Difusión e interiorización criterios	12211,000	259667,00	-3,747	,000	0,0187 PEQUEÑO
C.5.3. Resultados claros, confidenciales y concisos	9817,000	205442,00	-4,571	,000	0,0310 PEQUEÑO
C. 6.1. Oferta feedback	9074,000	205952,00	-5,132	,000	0,0384 PEQUEÑO
C. 6.2. Pautas proacción	8813,000	211379,00	-5,450	,000	0,0421 PEQUEÑO
C.7.1. Participación discente	13826,500	257777,50	-2,591	,010	0,0090 MUY PEQUEÑO

Nota: Las celdas sombreadas señalan aquellos resultados en los que se producen diferencias significativas dentro del intervalo de confianza establecido ($p < .05$). Las dimensiones comparadas son D1. (Planificación estratégica de la evaluación), D2 (Claridad de las condiciones y criterios de la evaluación), D7 (Participación discente en la evaluación), D5-II (Regulación del aprendizaje por medio de la evaluación redefinida y común) y TOTAL.

^a La información relativa a los ítems C.2.1. y C.7.1. replica los resultados presentados en las dimensiones 2 y 7, ya que representan los únicos componentes de las mismas.

Continuando la tendencia de respuesta descrita para las dimensiones, en los indicadores las diferencias significativas se producen en favor del profesorado (ver tabla 154). Mientras que este colectivo estima un nivel de cumplimiento *adecuado* en los seis elementos comparados, la realización percibida por el alumnado no alcanza la media teórica en ningún caso.

Tabla 154

Diferencias significativas en el contraste de los promedios de realización de los indicadores en función de la FUENTE: estadísticos de grupo

Factores	Fuente	<i>n</i>	<i>M</i>	<i>s</i>	<i>SEM</i>
C.1.1. Sistema evaluación continua	Docente	50	3,44	,929	,131
	Discente	703	2,94	1,092	,041
C.2.1. Difusión e interiorización criterios	Docente	50	3,18	,896	,127
	Discente	703	2,62	1,131	,043
C.5.3. Resultados claros, confidenciales y concisos	Docente	50	3,04	,880	,124
	Discente	625	2,39	1,029	,041
C. 6.1. Oferta feedback	Docente	50	3,12	,895	,127
	Discente	627	2,27	1,180	,047
C. 6.2. Pautas proacción	Docente	50	3,14	,756	,107
	Discente	636	2,27	1,187	,047
C.7.1. Participación discente	Docente	50	2,34	1,002	,142
	Discente	698	1,99	1,067	,040

Breve avance de la discusión de resultados...

Estudiantes y profesorado son los principales agentes implicados en la evaluación del aprendizaje, aunque dan cuenta de vivencias y percepciones diferentes en torno a la misma (V. Álvarez et al., 2002). Confirmando una tendencia de respuesta ampliamente constatada desde la investigación específica (Gutiérrez et al., 2011; Mas, 2012; Tejedor, 1999), el alumnado consultado atribuye mayor relevancia y menor grado de realización que el profesorado a los indicadores analizados. Salvada dicha tendencia, ambos sectores de la muestra coinciden al determinar los espacios de mayor y menor prioridad en la praxis evaluativa.

Tanto para estudiantes como para profesorado, la participación discente en la evaluación representa el factor menos importante y el menos realizado. Esta conclusión no es consistente con la amplia literatura que respalda los beneficios de la evaluación participativa desde diversas perspectivas (Boud & Falchikov, 2006; Falchikov & Goldfinch, 2000; A. Pérez et al., 2009). Sin embargo, confirma los hallazgos de otros estudios que se han aproximado a su exploración empírica, señalando una escasa incidencia de la misma (Gutiérrez et al., 2011, 2013; Rodríguez et al., 2012; Ruiz-Gallardo et al., 2013; Smitter, 2006).

En el extremo opuesto, alumnado y profesorado coinciden al destacar los procesos de retroinformación y proacción como el más relevante de los factores asociados a la competencia evaluativa. Este acuerdo respalda una conclusión ampliamente reportada desde la literatura específica (Black & Wiliam, 1998b; J. Chappuis et al., 2012; Hattie & Timperley, 2007; Jessop, 2013; Nicol, 2007, 2009; Yang & Carless, 2013). Sin embargo, al valorar el grado de realización del feedback y las pautas de mejora, estudiantes y profesorado ponen de manifiesto visiones diferentes, siendo el primer colectivo el que ofrece una postura más crítica. Estas discrepancias han sido señaladas en la misma dirección en otros estudios similares (Arribas et al., 2010; Gutiérrez et al., 2011, 2013; L. F. Martínez et al., 2012; Martínez-Mínguez & Flores, 2014). Una posible interpretación que se constata en torno a las mismas, es la desigual interpretación que sostienen sobre la naturaleza del feedback docentes y discentes (J. Chappuis, 2005, 2012; Higgins et al., 2002; Jonsson, 2013; Price et al., 2010).

5. Resultados relativos al Objetivo H. Formación permanente en evaluación

Este último objetivo del marco empírico centra la atención en la formación permanente del profesorado sobre la evaluación del aprendizaje. Plantea dos subpropósitos complementarios para realizar un doble análisis que atienda a las características de la formación cursada y a las demandas y necesidades formativas que manifiesta el profesorado. En ambos casos la información manejada procede de las respuestas de la muestra docente a dos cuestiones abiertas o semi-abiertas integradas en el cuestionario.

5.1 Resultados Objetivo H1. Describir los principales rasgos de las actividades formativas realizadas por el profesorado en los últimos años con relación a la evaluación del aprendizaje.

Atendiendo a las respuestas registradas, sólo el 23,52% de la muestra docente ha llevado a cabo alguna actividad formativa relacionada con la evaluación del aprendizaje en los dos últimos años. Como refleja la figura 46, la valoración de las mismas es positiva, sin que se registren respuestas que la califiquen como poco o nada útil o satisfactoria.

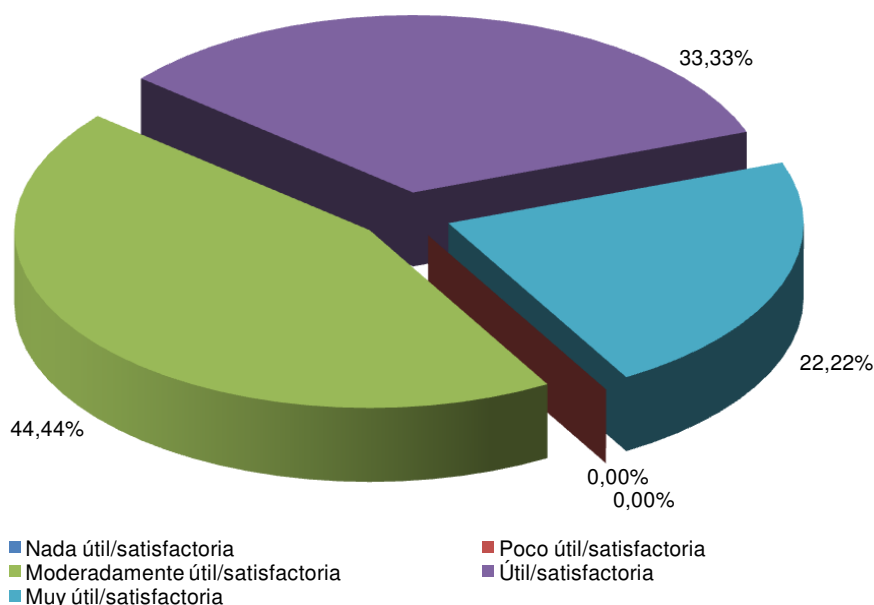


Figura 46. Satisfacción de la muestra docente con la formación permanente evaluativa desarrollada en los dos últimos años.

Las/os docentes que han llevado a cabo actividades formativas sobre la evaluación del aprendizaje señalan que su participación en las mismas ha sido voluntaria en el 90,9% de los casos. El porcentaje restante, refiere una combinación de motivaciones (no ha sido totalmente voluntaria, pero tampoco expresamente obligatoria). La duración media de dicha formación se sitúa levemente por debajo de las 38 horas ($M = 37,78$) aunque existen grandes diferencias entre los datos recopilados ($s = 34,60$).

Preguntado sobre las modalidades formativas, el profesorado del grupo con formación permanente reciente ha realizado un total de 15 aportaciones (ver tabla 155). Su registro, análisis y posterior categorización revela que la *autoformación* constituye la vía más extendida ($f = 7$; 46,66%). Dentro de esta categoría las/os docentes incluyen la lectura de manuales, artículos y documentos, los intercambios con colegas o la asistencia a conferencias y jornadas. El clásico *formato curso*, caracterizado por la presencialidad y la dirección de un agente externo experto, se señala en 3 ocasiones (el 20%), mientras que sólo se constata una única mención a su equivalente on-line (*on-line con tutorización*). Por último, otra sección del grupo describe la realización de módulos específicos sobre la evaluación del aprendizaje universitario dentro de *formación de postgrado más amplia*, al nivel de máster. Estos transcurren de forma presencial ($f = 3$) o en línea ($f = 1$).

Tabla 155
Principales modalidades empleadas en la formación permanente evaluativa del profesorado de la muestra

MODALIDADES DE FORMACIÓN REALIZADAS	<i>f</i>	%
Autoformación (lectura, intercambios con colegas, asistencia a jornadas, etc.)	7	46,66
Vinculada a formación de postgrado (máster)	4	26,66
⇒ Presencial	3	20,00
⇒ On-line	1	6,66
Presencial con dirección experta (formato curso)	3	20,00
On-line con tutorización	1	6,66
TOTAL DE MODALIDADES REPORTADAS	15	100

Como resume la tabla 156, la formación llevada a cabo gira en torno a cinco núcleos temáticos relacionados, con una incidencia muy similar en el banco de respuestas. La *evaluación de competencias en el marco del EEES* representa el tema más reiterado ($f = 5$; 27,7%). Aunque no incluyan ninguna alusión expresa a las competencias ni al proceso de convergencia, la formación sobre *herramientas e instrumentos de evaluación* (16,66% de

las aportaciones) y la aplicación de las TICs a la evaluación (22,22%) podrían estar sugiriendo focos de atención muy próximos a ese primer tema. Desde una perspectiva teórica más amplia, el profesorado describe acciones formativas sobre el *planteamiento general de la evaluación*, así como sobre su valor didáctico pedagógico (*evaluación para el aprendizaje*).

Tabla 156

Principales temas abordados en la formación permanente evaluativa del profesorado de la muestra

TEMAS ABORDADOS	f	%
Evaluación de competencias en el EEES	5	27,77
Planteamiento teórico general de la evaluación (modelos, principios, etc.)	4	22,22
Evaluación y TICs	4	22,22
Instrumentos de evaluación	3	16,66
Evaluación para el aprendizaje	2	11,11
TOTAL DE TEMAS REPORTADOS	18	100

Breve avance de la discusión de resultados...

En contraste con las necesidades declaradas (objetivo H2), los datos recopilados de la muestra indican que se efectúa poca formación permanente en materia evaluativa. El escaso interés que suscita la evaluación como foco formativo ha sido previamente sugerido desde perspectivas diferentes (V. Álvarez et al., 2011; JCSEE, 2007; Mas, 2012) y viene a agravar las dificultades asociadas a la falta de capacitación docente específica en el acceso a la enseñanza universitaria, ampliamente tratadas en la literatura (Erwin, 2003; Monereo & Castelló, 2009a; Valcárcel, 2003). Esta circunstancia podría predisponer al profesorado a incurrir con mayor facilidad en la reproducción de patrones evaluativos clásicos (B. Álvarez et al., 2007; Palacios & López, 2013; Tejedor, 1999).

En cuanto a las actividades realizadas, en la muestra se reafirma el voluntarismo apuntado por Zabalza (2002) o Mas (2012), siendo la autoformación la más recurrida de las alternativas para la capacitación o actualización permanente en materia evaluativa (lecturas, intercambios con colegas, etc.). Como señala Montero (2006), la influencia del EEES se percibe claramente en las temáticas abordadas desde la formación. Aunque el profesorado manifiesta un grado de satisfacción razonable con las actividades desarrolladas, no han sido recabados datos sobre su transferencia al ejercicio docente o el aprendizaje. Este último nivel de análisis representa un foco de reflexión relevante con una presencia cada vez mayor en la investigación próxima (Feixas, Fernández, Lagos, Quesada, & Sabaté, 2013; Monereo, 2013).

5.2 Resultados Objetivo H2. Identificar las principales necesidades formativas manifestadas por el profesorado con relación al desarrollo de su competencia evaluativa.

Según las respuestas directas registradas en los cuestionarios el 82,35% del profesorado consultado reconoce necesidades formativas relacionadas con la evaluación del aprendizaje de las/os estudiantes. El análisis y categorización de las contribuciones de ese sector mayoritario de la muestra desemboca en una serie de orientaciones sobre los temas a los que debería de dirigirse la formación y sobre las modalidades organizativas predilectas para su desarrollo.

Concentradas en la tabla 157, las aportaciones del profesorado sobre los formatos o modalidades formativas giran en torno a cinco núcleos. El clásico formato de curso, generalmente centrado en un programa definido y desarrollado en régimen presencial y con asesoramiento experto, no se encuentra entre ellos. En su lugar, se sugieren otro tipo de alternativas más centradas en el análisis y difusión de prácticas y situaciones reales:

Tabla 157
Principales modalidades de formación permanente sugeridas por el profesorado para la capacitación evaluativa

MODALIDADES DE FORMACIÓN SUGERIDAS	f	%
Difusión de buenas prácticas e intercambio de experiencias	7	35,00
⇒ Difusión de buenas prácticas e intercambio de experiencias (discusión e intercambio con colegas del propio centro)	5	25,00
⇒ Difusión de buenas prácticas e intercambio de experiencias en redes (discusión e intercambio con colegas de otras instituciones y lugares del mundo)	2	10,00
Asesoramiento experto <i>in situ</i>	5	25,00
Metaevaluación o revisión de la evaluación	4	20,00
⇒ Autorrevisión del sistema de evaluación guiada por criterios e indicadores	3	15,00
⇒ Revisión guiada del propio sistema de evaluación con participación experta	1	5,00
Investigación autónoma sobre la propia práctica	2	10,00
Análisis de casos reales	2	10,00
TOTAL DE SUGERENCIAS RECABADAS	15	100

En correspondencia con uno de los principales apoyos reseñados por el profesorado para el ejercicio de la evaluación (objetivo E3), el *intercambio de experiencias y la difusión de buenas prácticas* representa la modalidad formativa más reseñada ($f = 7$; 35%). Este

intercambio puede realizarse dentro de la propia institución (25%) o fuera de la misma, con ayuda de los grupos o redes profesionales (10%).

El segundo formato con mayor presencia entre las sugerencias recabadas es el *asesoramiento experto in situ* (25%). En esta modalidad el profesorado describe su interés por contrastar con un/a asesor/a especializada/o las dudas y los problemas que le genera la evaluación del aprendizaje en la práctica (aclaraciones y soluciones particulares, ajustadas a su situación particular); se ha denominado *in situ* porque en las reseñas existen referencias directas a la consultoría en el lugar de trabajo.

Ligado probablemente a los dos núcleos anteriores, entre las sugerencias docentes también se cuentan el *análisis de casos prácticos reales* y *los procesos de investigación-acción sobre la propia práctica evaluativa*.

La revisión de los sistemas de evaluación configurados o la *metaevaluación* representa otro de los formatos de formación permanente sugeridos por el profesorado para el desarrollo de su competencia evaluativa. Para su desarrollo se apela a actividades de autoevaluación a partir de un sistema de criterios e indicadores facilitados externamente (15%) y, en menor medida, a la evaluación externa a cargo de agentes expertos (5%).

Al detallar los temas en los que debería centrarse la formación permanente en materia evaluativa, las respuestas docentes desprenden 56 aportaciones. Realizadas de forma escueta y sintética, las temáticas sugeridas han sido agrupadas de forma conservadora (ver tabla 158), aunque entre ellas se presumen relaciones y posibles solapamientos:

Tabla 158
Principales temas de formación permanente sugeridos por el profesorado para la capacitación evaluativa

TEMAS DE FORMACIÓN SUGERIDOS	f	%
Evaluación de competencias genéricas y específicas	10	17,85
Evaluación viable/sostenible en grupos grandes	10	17,85
⇒ Instrumentos y fórmulas de feedback rápido en grupos numerosos	4	7,14
⇒ Sistemas de evaluación continua en grupos grandes	2	3,57
⇒ Métodos y técnicas de evaluación en grupos grandes (tipos de prueba, herramientas, etc.)	4	7,14
Métodos y técnicas de evaluación	7	12,5
⇒ Distintos tipos de prueba, herramientas, etc.	5	8,92
⇒ Métodos y técnicas de evaluación basados en TICs	2	3,57
Formación básica y actualización (fases, modalidades, planteamiento general, etc.)	6	10,71

TEMAS DE FORMACIÓN SUGERIDOS	f	%
Principios evaluativos	5	8,92
⇒ Principios técnicos (validez, fiabilidad.)	4	7,14
⇒ Principios éticos (justicia, transparencia, etc.)	1	1,78
Participación del alumnado en la evaluación (procedimientos y ponderación)	5	8,92
Orientaciones para el uso pedagógico/ estratégico de la evaluación	4	7,14
Planificación de la evaluación	4	7,14
⇒ Formulación de criterios e indicadores de evaluación	1	1,78
⇒ Selección de tareas y evidencias de aprendizaje significativas y adecuadas	3	5,35
Atención a la diversidad	4	7,14
⇒ Atención a la diversidad o adecuación de la evaluación en grupos grandes	3	5,35
⇒ Diagnóstico e intervención en dificultades de aprendizaje	1	1,78
Evaluación del aprendizaje colaborativo y trabajo en grupo	1	1,78
TOTAL DE TEMAS SUGERIDOS	56	100

En conexión directa con las mayores dificultades reseñadas por el profesorado para el ejercicio de sus funciones evaluativas (objetivo E3), la *evaluación de competencias* genéricas y específicas y la *viabilidad y sostenibilidad* de los sistemas de evaluación adoptados destacan como temas de formación prioritarios (17,85% de las aportaciones, respectivamente). La inquietud por la evaluación de competencias se expresa en la demanda de pautas y orientaciones para abordar su valoración. En cuanto a la viabilidad, las/os docentes solicitan formación para aprender a gestionar la sobrecarga asociada a la evaluación (mecanismos ágiles de feedback aplicables en grupos numerosos, formatos económicos de evaluación continua, etc.).

En aparente relación con los dos temas anteriores, las *técnicas y métodos de evaluación* representan otro foco de interés en lo que se desprende de las sugerencias recabadas ($f = 7$; 12,5%). Las respuestas del profesorado expresan el deseo de conocer nuevos medios que incrementen su repertorio metodológico y favorezcan una elección más consciente y adecuada en cada caso. Algunas de las respuestas hacen referencia directa a los recursos evaluativos vinculados a las TICs ($f = 2$; 3,57%). La demanda de capacitación en *planificación evaluativa* ($f = 4$; 7,14%), centrada en la formulación de indicadores y criterios de evaluación y en la selección de tareas y evidencias de aprendizaje significativas y adecuadas, también parece relacionada con el interés por la evaluación viable de competencias.

El 10,71% de las aportaciones registradas proponen la organización de acciones que procuren al profesorado una *formación básica sobre evaluación*. Las/os participantes

describen en este bloque una actualización general sobre el tema, que revise aspectos fundamentales como las modalidades evaluativas, las fases del proceso o conceptos básicos asociados. Estrechamente relacionado, el tema de los *principios evaluativos* de referencia se presenta como sugerencia en algo menos del 9% de las aportaciones registradas. La muestra apela a la profundización en fundamentos éticos o en los principios técnicos de fiabilidad y validez, todos ellos contemplados en el factor Metaevaluación del cuestionario respondido. Esta coincidencia refuerza la sospecha de posible transformación de la muestra, y se reitera en la propuesta de dos temáticas más, explícitamente abordadas en la herramienta de autoevaluación: las *fórmulas de evaluación participativa* ($f = 5$; 8,92%) y las *orientaciones para el uso estratégico* de la evaluación ($f = 4$; 7,14%).

Las dificultades expresadas por el profesorado de la muestra para el ejercicio de la evaluación (objetivo E3) tienen reflejo en dos núcleos más de interés formativo. Así, la *atención a la diversidad en la evaluación* se sugiere como objeto de formación en el 7,14% de las respuestas recabadas (adecuación en grupos grandes y diagnóstico y tratamiento de dificultades), junto a la valoración del *aprendizaje cooperativo y el trabajo en grupo* (1,78%).

Breve avance de la discusión de resultados...

La generalización de una formación didáctica específica, que capacite para el ejercicio de la docencia universitaria, representa una demanda constante (V. Álvarez et al., 2011; Delbanco, 2012; Monereo, 2013, 2014a), especialmente en las dos últimas décadas, a raíz de los nuevos retos asociados a la implantación del proceso de convergencia (M. González & Raposo, 2009; Ibarra & Rodríguez, 2010). Al tratar sobre la capacitación evaluativa en particular, se produce esa misma tendencia. La formación del profesorado en este terreno representa una sugerencia de mejora recurrente en los estudios que analizan la praxis evaluativa (V. Álvarez et al., 2002; Caballero, 2013; B. Gargallo et al., 2010; Palacios & López, 2013). Así se advierte también en la muestra analizada, donde una amplia mayoría del profesorado hace suya la necesidad de apoyo formativo en este espacio de la actividad docente. Los temas y modalidades precisos sugeridos para su desarrollo muestran una alta correspondencia con las orientaciones y sugerencias que al respecto arroja la literatura.

En lo que respecta a las modalidades formativas, la literatura e investigación específica toma distancia del clásico formato de transmisión unidireccional y sugiere el trabajo conjunto, el diálogo y el intercambio en torno a los problemas e inquietudes reales que genera la praxis evaluativa (Blanch & Corcelles, 2014; Capllonch & Buscà, 2012; Mateo, 2012; Torra et al., 2013). Se apuesta por la formación in situ, apegada y desarrollada en el contexto real de aplicación, admitiendo que en ella pueda tener cabida la participación de agentes expertos en calidad de asesores (Feixas et al., 2013; Mauri et al., 2013; Southwell & Morgan, 2009). Así parece entenderlo también el profesorado de la muestra, que sugiere experiencias de intercambio, asesoramiento o investigación enraizadas en su

propio trabajo y en los dilemas evaluativos que en el mismo se plantean. Por la implicación, la participación y el perspectivismo (Rodrigo, 1993) que provocan, este tipo de experiencias parecen especialmente propicias para contribuir al cambio en las concepciones evaluativas profundas.

En la propuesta de posibles temas de formación evaluativa también se aprecia el mismo nivel de correspondencia. En coherencia con los grandes ejes de mejora que sugieren el Assessment Resource at Hong Kong University (AR@HKU, 2009d) y el Centre for the Study of Higher Education (CSHE, 2002), la evaluación viable o sostenible en grupos numerosos destaca como una de las inquietudes prioritarias. Las referencias al feedback están especialmente presentes en esta demanda, en correspondencia con las necesidades de las que se hace eco la literatura (Carless et al., 2011; Hounsell, 2007; Lu & Law, 2012; Nicol et al., 2014). Admitida la particular relevancia de este espacio de la evaluación (objetivo E1), el profesorado solicita formación que le capacite para proporcionar retroalimentación efectiva en las condiciones reales de trabajo.

La evaluación de competencias y los instrumentos y herramientas aplicables a la misma también figura en el listado de cuestiones a tratar, a pesar de ser uno de los focos más representados en la formación llevada a cabo (objetivo H1). Probablemente por influencia de los planteamientos pedagógicos del EEES (Montero, 2006) o por la complejidad natural que entraña su ejercicio (L. F. Martínez et al., 2012), el profesorado manifiesta interés por profundizar en las implicaciones evaluativas del enfoque formativo basado en competencias. Probablemente ligada a las necesidades que plantea su valoración, también se constata una demanda de formación sobre instrumentos y herramientas de evaluación. A ese respecto es importante considerar los riesgos de una preparación excesivamente centrada en la dimensión técnica o instrumental, insuficiente para promover el cambio en las concepciones evaluativas profundas por sí sola (Bordas & Cabrera, 2001; Cano & Ion, 2012; Knight, 2005; Verdejo, 2008).

6. El capítulo a grandes rasgos

A lo largo del octavo capítulo se ha llevado a cabo la exposición de los resultados obtenidos en los cuatro objetivos vinculados al marco empírico de la investigación. A continuación se muestran resumidas las principales conclusiones evidenciadas en cada uno de ellos, antes de proceder a su discusión pormenorizada en el próximo capítulo.

OBJETIVO E. Realizar una autoevaluación de la competencia evaluativa del profesorado dedicado a la formación básica de maestras/os de Educación Infantil y/o Primaria en el contexto institucional de la UD y BAM.

El análisis de este primer objetivo ha sido abordado por medio de tres sub-propósitos:

OBJETIVO E1. Identificar los factores e indicadores de la evaluación del aprendizaje en los que mayor y menor dificultad expresa el profesorado, atendiendo al nivel de relevancia y realización que les atribuyen.

En términos generales el profesorado se considera competente en materia evaluativa, aunque todos los factores analizados se sitúan en el área de Concentración de la rejilla AIR.

Según revelan las respuestas docentes, los factores más importantes representan también los más llevados a cabo. La Regulación del aprendizaje y la enseñanza por medio de la evaluación, la Planificación estratégica de la misma o la Puesta en práctica de los dispositivos diseñados, se consideran aspectos de la evaluación especialmente relevantes ($M \geq 4$), en los que la autopercepción de cumplimiento supera levemente la media teórica ($3,22 \leq M \leq 3,34$; *cumplimiento adecuado*).

Los tres factores restantes se sitúan en el polo opuesto por distintas circunstancias. Los Ajustes en la evaluación y la participación discente representan para la muestra docente el menos importante de los espacios valorados, mientras que la Metaevaluación destaca como el menos llevado a cabo. Próximo a ellos, la Adecuación y precisión de la evaluación representa el segundo factor menos importante en opinión del profesorado y también el segundo menos reflejado en la práctica.

OBJETIVO E2. Identificar relaciones entre distintas variables académicas, socio-demográficas o profesionales y la competencia evaluativa autopercebida.

Las valoraciones de importancia y realización del profesorado para los seis factores ligados a su competencia evaluativa han sido analizadas a la luz de 17 variables de tipo

académico, profesional o socio-demográfico. Como resume la tabla 159, las diferencias de promedios advertidas han resultado estadísticamente significativas (IC = 95%; $p < .05$) en 10 de ellas, siendo de tamaño grande el efecto medio estimado para las mismas ($\eta^2 = 0,1523$).

Tabla 159

Competencia evaluativa autopercebida y variables académicas, socio-demográficas o profesionales: síntesis de resultados (objetivo E2)

ANÁLISIS	RESULTADOS CONCLUIDOS
1. SEXO	• IMP. Factor 3, donde las mujeres ofrecen valoraciones superiores (TE MEDIO). • IMP. Total, donde las mujeres vuelven a registrar promedios superiores (TE MEDIO).
2. EDAD	• Correlación positiva y baja entre la edad y REA. Factor 2.
3. EXPERIENCIA DOCENTE	• No existen correlaciones significativas entre la experiencia docente y las valoraciones de importancia y realización. • REA. Factor 2, diferencias significativas entre el grupo de mayor experiencia y el medio (TE GRANDE).
4. FORMACIÓN PREVIA	• Ocho grupos profesionales: ✓ REA. Total, donde el profesorado con formación previa en Filología obtiene menores promedios que el titulado en Filosofía (TE GRANDE). ✓ REA. Factor 1, donde el profesorado con formación previa en Filología obtiene menores promedios que el titulado en Pedagogía o Psicopedagogía (TE GRANDE). ✓ REA Factor 4, donde se producen diferencias entre cinco de los ocho grupos profesionales (TE GRANDE). • Docentes con formación profesional previa en Pedagogía o Psicopedagogía y resto del profesorado: ✓ REA. Factor 1, donde el profesorado con formación previa en Pedagogía o Psicopedagogía obtiene mayores promedios (TE GRANDE). ✓ IMP. Factor 6, donde el profesorado con formación previa en Pedagogía o Psicopedagogía obtiene mayores promedios (TE MEDIO). • Docentes con formación profesional previa en Psicología, Pedagogía o Psicopedagogía y resto del profesorado: ✓ IMP. Total (TE MEDIO) y REA. Total (TE GRANDE), donde el profesorado con formación previa en Psicología, Pedagogía o Psicopedagogía obtiene promedios superiores. ✓ IMP. Factor 1 (TE MEDIO) y REA. Factor 1 (TE GRANDE), donde el profesorado con formación previa en Psicología, Pedagogía o Psicopedagogía obtiene promedios superiores. ✓ IMP. Factor 4 (TE MEDIO) y REA. Factor 4 (TE GRANDE), donde el profesorado con formación previa en Psicología, Pedagogía o Psicopedagogía obtiene promedios superiores. • Docentes con formación profesional previa en Filología y resto del profesorado: ✓ REA. Total, donde el profesorado con formación previa en Filología obtiene menores promedios (TE GRANDE). ✓ REA. Factor 1, donde el profesorado con formación previa en Filología obtiene menores promedios (TE GRANDE). ✓ REA Factor 4, donde el profesorado con formación previa en Filología obtiene menores promedios (TE GRANDE). • No se producen diferencias estadísticamente significativas entre las valoraciones de docentes con formación profesional en ciencias naturales o sociales.

ANÁLISIS	RESULTADOS CONCLUIDOS
5. MÁSTER	No existen diferencias estadísticamente significativas.
6. DOCTORADO	
7. DEDICACIÓN	
8. SIMULTANEIDAD	• IMP. Factor 1, donde los promedios superiores se producen en el grupo con dedicación exclusiva a la docencia universitaria (TE MEDIO).
9. CENTRO	• REA. Factor 3, donde el profesorado de la UD muestra promedios superiores (TE MEDIO). • IMP. Factor 6, donde el profesorado de la UD muestra promedios superiores (TE MEDIO).
10. GRADO	No existen diferencias estadísticamente significativas.
11. CURSO	• REA. Factor 1, donde el profesorado que ha participado revisando su competencia evaluativa en módulos/asignaturas de primer curso, produce mayores promedios que las/os que lo han hecho desde segundo o tercero (TE MEDIO). • IMP. Factor 5, donde el profesorado que ha participado revisando su competencia evaluativa en módulos/asignaturas de primer curso, produce mayores promedios que las/os que lo han hecho desde segundo o tercero (TE GRANDE).
12. MÓDULO	No existen diferencias estadísticamente significativas.
13. MATERIA	
14. UNIDAD	
15. COMPETENCIA EVALUATIVA AUTOPERCIBIDA	• REA. Factor 1, donde el profesorado que declara poca competencia muestra promedios inferiores respecto al grupo que se considera competente o muy competente (TE MEDIO). • REA. Factor 5, donde el profesorado que declara poca competencia muestra promedios inferiores respecto al grupo que se considera competente o muy competente (TE MEDIO).
16. FORMACIÓN EVALUATIVA REALIZADA	• REA. TOTAL, donde el profesorado que afirma haber realizado formación evaluativa presenta promedios superiores (TE MEDIO). • IMP. y REA. Factor 4, donde el profesorado que afirma haber realizado formación evaluativa presenta promedios superiores (TE MEDIO).
17. NECESIDAD DE FORMACIÓN	• IMP. Factor 3, donde el profesorado que manifiesta necesidad formativa presenta promedios superiores (TE MEDIO).

Nota: Las conclusiones resumidas hacen referencia a la subescala de importancia (IMP.) y a la de realización (REA.). Los factores (F) se presentan en el mismo orden que han sido definidos: F1. Planificación estratégica de la evaluación; F2. Regulación del aprendizaje y de la enseñanza por medio de la evaluación; F3. Metaevaluación; F4. Ajustes en la evaluación y participación discente; F5. Adecuación y precisión; F6. Puesta en práctica de la evaluación.

Según ha sido descrito al presentar los resultados, resultan especialmente destacables las diferencias advertidas en función de la formación previa y permanente del profesorado. Así se concluye por la cantidad y el tamaño de las diferencias de apreciación que generan dentro de la muestra, así como por su posible relación con otras de las conclusiones significativas detectadas en la exploración de la perspectiva discente (objetivo F2).

OBJETIVO E3. Identificar las principales fortalezas, debilidades y dificultades para el ejercicio de la evaluación del aprendizaje desde el punto de vista docente.

En las respuestas producidas el profesorado de la muestra presta mayor atención a los aspectos que dificultan su praxis evaluativa que a los factores que la apoyan o facilitan.

Los mayores obstáculos percibidos en la evaluación del aprendizaje son la excesiva carga de trabajo, la atención individualizada en la evaluación y la falta de formación.

- La *sobrecarga* representa la dificultad más reiterada e incluye cinco aspectos con incidencia en el volumen de la tarea docente y su temporalidad (número de estudiantes, sobrecarga de funciones, mayor exigencia de trabajo docente asociada a los sistemas de evaluación formativa y continua, condiciones prácticas de la evaluación y cambios en la adjudicación o encargo docente de un curso a otro).
- Ligado a ello, la *atención individualizada* en la evaluación representa otro factor difícil de sobrellevar a juicio docente. En ella se incluyen las limitaciones prácticas para hacer frente a dos cometidos: el seguimiento individualizado y la comisión de feedback personalizado al alumnado, y la adecuación de los dispositivos evaluativos a las necesidades particulares dentro del grupo.
- La *falta de formación en evaluación* representa el tercer obstáculo más reseñado por el profesorado. Hace referencia a la complejidad técnica que entraña la evaluación de competencias y a una cierta inseguridad y falta de preparación con respecto a los procedimientos evaluativos (principios, técnicas, instrumentación adecuada, etc.).

En cuanto a los apoyos en el ejercicio de la evaluación, *el intercambio entre colegas* y *la planificación rigurosa* representan las ayudas más reseñadas dentro de la muestra. Le siguen la *formación* y *el asesoramiento externo en materia evaluativa*, así como la *revisión y ajuste de los dispositivos configurados* a partir de la experiencia.

OBJETIVO F. Realizar un diagnóstico de la competencia evaluativa del profesorado dedicado a la formación básica de maestras/os de Educación Infantil y/o Primaria en la UD y BAM desde la perspectiva de las/os estudiantes.

Replicando el esquema anterior, las conclusiones de este segundo objetivo se ordenan de acuerdo a tres sub-propósitos interconectados:

OBJETIVO F1. Identificar los factores e indicadores de la evaluación del aprendizaje en los que mayor y menor dificultad acusa el alumnado, atendiendo al nivel de relevancia y realización que les atribuyen.

Los datos recopilados de la muestra discente, sitúan la competencia evaluativa del profesorado en el área de Concentración de la rejilla AIR. Los factores e indicadores asociados a la misma se consideran entre *relevantes* y *muy relevantes* ($M = 4,29$; $s = 0,37$), pero su promedio de realización percibida se sitúa por debajo de la media teórica ($M = 2,82$; $s = 0,61$). No obstante, este patrón de respuesta no se traslada de igual manera a los seis factores valorados.

Desde el punto de vista del alumnado, la Regulación del aprendizaje y su Participación en la evaluación representan los factores menos llevados a cabo en la práctica. Compartida esta premisa, su alcance parece diferente en lo que se desprende de las valoraciones registradas:

- Para las/os estudiantes su Participación en la evaluación representa el menos relevante de todos los factores y, además, el menos realizado.
- Mayor es el desajuste apreciado en la Regulación del aprendizaje, que destaca como el factor de la competencia evaluativa docente más importante y como el segundo menos llevado a cabo.

Frente a ellos, el alumnado considera que los Ajustes en la evaluación representan el factor mejor desempeñado en la práctica, hasta el punto de ubicarlo en la zona de Mantenimiento de la rejilla AIR. Aunque con un promedio de realización algo menor, en su opinión el Control de las condiciones de la evaluación también se encuentra en mejor situación que el resto de factores.

Las valoraciones discentes reflejan una situación similar para los dos factores restantes. La Planificación del sistema evaluativo y la Adecuación y precisión de la evaluación se consideran factores entre *relevantes* y *muy relevantes* ($M = 4,28$; $M = 4,40$), con un nivel de cumplimiento que justo alcanza la media teórica ($M = 3,03$, *adecuado*). No obstante, el elevado índice de dispersión constatado en la sub-escala de realización en ambos casos ($s = ,73746$; $s = 1,02028$, respectivamente) incita a realizar una interpretación conservadora de estos resultados, en previsión de discrepancias y vivencias diferentes dentro de la muestra.

OBJETIVO F2. Identificar relaciones entre distintas variables académicas, profesionales o socio-demográficas y la competencia evaluativa percibida por el alumnado.

Los análisis practicados han puesto de manifiesto diferencias significativas dentro del intervalo de confianza establecido ($IC = 95\%$; $p < .05$) en 14 de las 15 variables académicas, profesionales o socio-demográficas consideradas (ver tabla 160). Debido probablemente al tamaño de la muestra ($n = 704$) la magnitud media del efecto estimado para las mismas es pequeña ($\eta^2 = 0,0483$).

Tabla 160

Competencia evaluativa percibida por el alumnado en función de variables académicas, socio-demográficas o profesionales: síntesis de resultados (objetivo F2)

ANÁLISIS	RESULTADOS CONCLUIDOS
1. SEXO	• IMP. Total, (TE PEQUEÑO) y REA. Total (TE MUY PEQUEÑO). • IMP. Factor 1 (TE PEQUEÑO) y REA. Factor 1 (TE MUY PEQUEÑO). • IMP. Factor 2 y REA. Factor 2 (TE PEQUEÑO). • IMP. Factor 4 e IMP. Factor 6 (TE PEQUEÑO). • IMP. Factor 3 e IMP. Factor 5 (TE MUY PEQUEÑO). • En las valoraciones de importancia las mujeres ofrecen mayores puntuaciones, mientras que en la realización son los varones los que puntúan más.
2. EDAD	• Se detectan varias relaciones entre la edad y las valoraciones discentes (significativas o muy significativas). ✓ Correlación muy baja y positiva entre edad y puntuaciones en IMP. Total, IMP. Factor 2, IMP. Factor 4 e IMP. Factor 5. ✓ Correlación muy baja y negativa entre edad y puntuaciones en REA. Total, REA Factor 1 y REA Factor 2. • Diferencias significativas en varias valoraciones en función del grupo de edad. La tendencia general apreciada coincide con la correlación anterior: en los grupos de mayor edad las valoraciones de relevancia son superiores e inferiores las de cumplimiento (TE PEQUEÑO en todas las diferencias). ✓ IMP. Total, donde el alumnado de menor edad (18 años o menos) presenta promedios inferiores que el de 20 años. ✓ REA. Factor 1, donde el alumnado de menor edad presenta promedios superiores que el de las/os estudiantes de 20 años y de 21 años o más. ✓ IMP. Factor 2, donde el alumnado de menor edad presenta promedios inferiores a los de los dos grupos de mayor edad en la muestra. ✓ REA. Factor 2, donde el alumnado de menor edad presenta promedios superiores que el de las/os estudiantes de 20 años. ✓ IMP. Factor 4, donde el alumnado de menor edad presenta promedios inferiores a los de los dos grupos de mayor edad en la muestra. ✓ IMP. Factor 5, donde el alumnado de menor edad presenta promedios inferiores a los del grupo de 20 años.

ANÁLISIS	RESULTADOS CONCLUIDOS
3. DEDICACIÓN A LA FORMACIÓN	• IMP. Total, donde el alumnado con dedicación exclusiva ofrece menores valoraciones que el que trabaja sin contrato (TE PEQUEÑO). • IMP. Factor 2, donde el alumnado con dedicación exclusiva ofrece menores valoraciones que el que trabaja sin contrato (TE PEQUEÑO). • IMP. Factor 3, donde el alumnado con dedicación exclusiva ofrece menores valoraciones que el que trabaja sin contrato (TE PEQUEÑO). • REA. Factor 4, donde el alumnado que trabaja con contrato ofrece menores valoraciones que el que trabaja sin tenerlo (TE MUY PEQUEÑO).
4. VÍA DE ACCESO	No existen diferencias estadísticamente significativas.
5. GRADO	• REA. Total, REA. Factor 1 y REA. Factor 4, donde el alumnado de Educación Infantil ofrece menores promedios que el de Educación Primaria (TE PEQUEÑO). • IMP. Total, IMP. Factor 1 e IMP. Factor 3, donde el alumnado de Educación Infantil ofrece mayores promedios que el de Educación Primaria (TE PEQUEÑO). • IMP. Factor 2, donde el alumnado de Educación Infantil ofrece mayores promedios que el de Educación Primaria (TE MUY PEQUEÑO).
6. CURSO	• IMP. Total, IMP. Factor 2, IMP. Factor 4 e IMP. Factor 5, donde el alumnado de tercer curso ofrece mayores promedios que el de primero (TE PEQUEÑO). • IMP. Factor 1, donde el alumnado de tercer curso ofrece mayores promedios que el de segundo (TE PEQUEÑO). • REA. Total, REA. Factor 1, REA. Factor 2, REA. Factor 4 y REA. Factor 5, donde el alumnado de primer curso ofrece promedios superiores a los registrados en segundo y en tercero (TE PEQUEÑO). • REA. Factor 3, donde el alumnado de primer curso ofrece promedios superiores a los registrados entre estudiantes de tercero (TE PEQUEÑO).
7. CENTRO	• REA. Factor 2 y REA. Factor 4, donde el alumnado de la UD concede menores puntuaciones que el matriculado en BAM (TE MUY PEQUEÑO). • IMP. Factor 1, donde el alumnado de la UD concede mayores puntuaciones que el matriculado en BAM (TE MUY PEQUEÑO).
8. UNIDAD	• IMP. Total e IMP. Factor 4, donde el alumnado de los módulos ofrece mayores promedios que el de asignaturas (TE MUY PEQUEÑO). • REA. Total, REA. Factor 1, REA. Factor 2, REA. Factor 3, REA. Factor 4, REA. Factor 5 y REA. Factor 6, donde el alumnado de los módulos ofrece menores promedios que el de asignaturas. En el total de la subescala TE MEDIO, y en el resto de resultados PEQUEÑO o MUY PEQUEÑO.
9. MATERIA	• IMP. Total y REA. Factor 3 (TE MEDIO). • REA. Total, REA. Factor 1, REA. Factor 2, REA. Factor 4, REA. Factor 5 y REA. Factor 6 (TE GRANDE). • IMP. Factor 1, IMP. Factor 2, IMP. Factor 3, IMP. Factor 4, IMP. Factor 5, IMP. Factor 6 (TE PEQUEÑO).
10. MÓDULO	• Diferencias en todas las valoraciones de la subescala de realizaciones y en la importancia atribuida a los factores 4 y 5. En todas ellas TE PEQUEÑO. ✓ REA. Total, donde el alumnado del módulo didáctico disciplinar puntúa por debajo de las/os estudiantes de mención y del módulo de formación básica. ✓ REA. Factor 1, donde las diferencias afectan a varios grupos. Las valoraciones discentes para el módulo DECA resultan inferiores que las efectuadas en los otros tres módulos. En el módulo de formación básica los promedios de respuesta son mayores que en el didáctico disciplinar. ✓ REA. Factor 2, donde el alumnado del módulo didáctico disciplinar presenta promedios inferiores que el de los módulos mención y formación básica. ✓ REA. Factor 3, donde el alumnado del módulo didáctico-disciplinar concede menores puntuaciones que el del módulo de formación básica.

ANÁLISIS	RESULTADOS CONCLUIDOS
10. MÓDULO	✓ REA. Factor 4, donde las diferencias afectan a varios grupos. El alumnado del módulo didáctico disciplinar registra promedios menores que en los tres módulos restantes. El alumnado de formación básica realiza valoraciones inferiores a las que efectúa el de DECA. ✓ REA. Factor 5, donde las diferencias afectan a varios grupos. El alumnado del módulo didáctico-disciplinar registra menores promedios de respuesta que el de formación básica y el de mención. También entre estudiantes de DECA y del módulo de formación básica, siendo estos últimos quienes mayores puntuaciones conceden. ✓ REA. Factor 6, donde los promedios registrados en el módulo de formación básica resultan superiores a los de DECA y a los del módulo didáctico disciplinar. ✓ IMP. Factor 4, donde los promedios de respuesta del alumnado de mención son superiores a los registrados en el módulo de formación básica. ✓ IMP. Factor 5, donde los promedios de respuesta del alumnado del módulo didáctico-disciplinar superan a los registrados en el de formación básica.
11. EXPERIENCIA PREVIA	• REA. Factor 1, donde las/os estudiantes con experiencia previa con el profesorado declaran un mayor grado de realización (TE MUY PEQUEÑO).
12. RESULTADOS PARCIALES	• REA. Total, REA. Factor 2, REA. Factor 4 y REA. Factor 5, donde las/os estudiantes con resultados parciales sobre su rendimiento conceden mayores puntuaciones (TE PEQUEÑO).
13. EXPECTATIVAS DE ÉXITO	• REA. Total, REA. Factor 1, REA. Factor 5 y REA. Factor 6, donde el alumnado con expectativas positivas presenta mayores promedios (TE PEQUEÑO). • REA. Factor 2, se producen diferencias significativas pero no es posible determinar entre qué grupos (TE MUY PEQUEÑO). • IMP. Factor 4, donde el grupo de estudiantes con dudas sobre su calificación ofrece mayores promedios de respuesta (TE PEQUEÑO).
14. COMPETENCIA DOCENTE PERCIBIDA	• Diferencias significativas en todas las valoraciones de realización y entre los tres grupos de competencia percibida. Los mayores promedios los presentan quienes atribuyen alta competencia y los más bajos quienes atribuyen poca competencia. ✓ REA. Factor 3 (TE MEDIO). ✓ REA. Factor 4 (TE PEQUEÑO). ✓ REA. Total, REA. Factor 1, REA. Factor 2, REA. Factor 5 y REA. Factor 6 (TE GRANDE). • Correlación positiva y muy significativa con todas las valoraciones de realización: ✓ En REA. Total, REA. Factor 2 y REA. Factor 5 la asociación es moderada. ✓ En REA. Factor 1, REA. Factor 3, REA. Factor 4 y REA. Factor 6 la asociación es pequeña. • Correlación significativa alta y positiva con IMP. Factor 2.
15. MODALIDAD CONTRASTE	• REA. Total, REA. Factor 4, REA. Factor 5 y REA. Factor 6, donde las/os estudiantes participantes en la modalidad abreviada conceden mayores puntuaciones (TE PEQUEÑO). • IMP. Factor 4 y REA. Factor 1, donde las/os estudiantes participantes en la modalidad abreviada conceden menores puntuaciones (TE PEQUEÑO).

Nota: Las conclusiones resumidas hacen referencia a la subescala de importancia (IMP.) y a la de realización (REA.). Los factores se presentan en el mismo orden que han sido definidos: F1. Participación del alumnado en la evaluación; F2. Planificación del sistema evaluativo; F3. Ajustes del sistema evaluativo; F4. Regulación del aprendizaje por medio de la evaluación; F5. Adecuación y precisión; F6. Control de las condiciones en las que se desarrolla la evaluación. La letra cursiva empleada sugiere una lectura cautelosa de los resultados asociados al factor sexto, por los indicios de ambigüedad evidenciados en su estructuración interna.

El análisis de este segundo objetivo ha puesto de manifiesto un gran número de relaciones e influencias entre factores socio-demográficos, profesionales y académico-institucionales y la percepción discente de la competencia evaluativa. Aunque altamente significativas, las diferencias halladas se caracterizan por ejercer un efecto pequeño o muy pequeño ($\eta^2 \leq 0,01$). Las únicas variables que aparecen asociadas a efectos de magnitud grande son la materia y el nivel de competencia docente atribuido al profesor/a de referencia en la valoración.

OBJETIVO F3. Identificar las principales fortalezas y debilidades en la evaluación del aprendizaje desde el punto de vista discente.

Las aportaciones de las/os estudiantes en la descripción de las fortalezas y debilidades asociadas a la evaluación contemplan varios aspectos incluidos en el cuestionario aplicado para la valoración de la competencia evaluativa, lo que sugiere que ha podido tener lugar una cierta transformación de la muestra (Manzano & Braña, 2005). Como el profesorado, el alumnado consultado sostiene una visión más bien negativa de la evaluación de su aprendizaje, con mayor presencia de dificultades o áreas de mejora que de fortalezas.

A juzgar por la cantidad de alusiones que suma, el feedback recibido representa la mayor dificultad asociada a la evaluación para la muestra discente. Las/os estudiantes describen una *retroalimentación deficiente*, tardía y basada en orientaciones generales e insuficientemente detalladas. Unido a ello, la *ausencia de resultados parciales* destaca como la segunda área de mejora más reiterada en las respuestas del alumnado. Estas conclusiones adoptan una especial significación si se considera que la Regulación del aprendizaje por medio de la evaluación constituye el factor de mayor relevancia para la muestra discente; este último hace referencia tanto a la comunicación de resultados, como a los procesos de retroalimentación y proacción a partir de los mismos.

Al señalar las fortalezas de la evaluación, el alumnado destaca como principales aspectos a mantener la continuidad del proceso, la variedad de técnicas e instrumentos que se ponen en juego en la misma y la claridad en la información recibida sobre su planteamiento y desarrollo.

OBJETIVO G. Contrastar el punto de vista que manifiestan docentes y discentes con respecto a la competencia evaluativa del profesorado vinculado a la formación básica de maestras/os en la UD y BAM, indagando tanto en la perspectiva teórica (importancia), como en la práctica (realización).

Examinadas por separado la visión que manifiestan sobre la competencia evaluativa estudiantes y profesorado (objetivos F y E), el tercer objetivo del marco empírico ha planteado un análisis comparativo. Este ha tenido lugar sobre los nueve indicadores comunes y directamente contrastables en ambos sectores de la muestra y las cuatro dimensiones en las que se ordenan.

Las respuestas obtenidas de ambas fuentes se ajustan a un mismo patrón: mientras que en la subescala de importancia es el alumnado quien concede las mayores puntuaciones, al estimar el grado de cumplimiento la tendencia se invierte. Salvada esta tónica general, ambos sectores de la muestra revelan en sus valoraciones una priorización común de las cuatro dimensiones comparables asociadas a la competencia evaluativa docente. Estudiantes y profesorado coinciden en situar todas ellas en la zona de Concentración de la representación AIR. Dentro del desajuste que denota dicha posición, también coinciden en determinar situaciones diferentes entre los aspectos valorados:

- La Regulación del aprendizaje representa para docentes y estudiantes la más importante de las dimensiones asociadas a la competencia evaluativa.
- En el extremo opuesto, la Participación discente en la evaluación representa para ambos colectivos el menos importante de los indicadores cuestionados y, a la vez, el menos llevado a cabo en la práctica.
- La dimensión relativa a la Planificación estratégica de la evaluación es la única que ambos colectivos coinciden en situar por encima de la media teórica; también representa la que mayor promedio de realización concentra según estudiantes y profesorado.
- Por último, la Claridad de las condiciones y criterios de la evaluación es juzgada por ambos sectores de la muestra como una dimensión importante, aunque con diferencias significativas en lo que refiere a su percepción de cumplimiento.

La tabla 161 ofrece un resumen de las diferencias estadísticamente significativas ($p < .05$) detectadas en el contraste de las valoraciones procedentes de ambas fuentes.

Tabla 161

Competencia evaluativa percibida por el alumnado en función de variables académicas, socio-demográficas o profesionales: síntesis de resultados (objetivo G)

ANÁLISIS	RESULTADOS CONCLUIDOS
1. FUENTE Y VALORACIONES IMPORTANCIA	• IMP. Dimensión 4 comparable, donde el alumnado presenta promedios superiores de respuesta (TE MUY PEQUEÑO). Éstas tienen lugar en dos de los tres indicadores que componen el factor: ✓ Ítem C.5.3. (<i>Comunicación de los resultados de aprendizaje</i>) (TE PEQUEÑO). ✓ Ítem C.6.2. (<i>Oferta de pautas de proacción para la mejora</i>) (TE PEQUEÑO).
2. FUENTE Y VALORACIONES REALIZACIÓN	• Diferencias significativas en todos los resultados de la subescala de realización. Tanto en el total como en las cuatro dimensiones, los menores promedios de respuesta los concentra el alumnado: ✓ REA. Dimensión 1 y Dimensión 7 (TE MUY PEQUEÑO). ✓ REA. Total, REA. Dimensión 2 y Dimensión 4 comparable (TE PEQUEÑO). • Estas diferencias se manifiestan en diferentes indicadores dentro de las dimensiones: ✓ En la Dimensión 1 las diferencias se producen en el ítem C.1.1. (<i>Continuidad de la evaluación</i>) (TE PEQUEÑO). ✓ En la Dimensión 2 las diferencias se producen en el ítem C.2.1. (<i>Claridad de las condiciones y criterios de la evaluación</i>) (único componente) (TE PEQUEÑO). ✓ En la Dimensión comparable 4 las diferencias se producen en los ítems C.5.3. (<i>Comunicación de los resultados de aprendizaje</i>), 6.1. (<i>Comisión de feedback</i>) y 6.2. (<i>Oferta de pautas de proacción para la mejora</i>) (TE PEQUEÑO). ✓ En la Dimensión 7 las diferencias se producen en el ítem C.7.1. (<i>Participación discente en la evaluación</i>) (único componente) (TE MUY PEQUEÑO).

Nota: Las conclusiones resumidas hacen referencia a la subescala de importancia (IMP.) y a la de realización (REA.). Las dimensiones comparables son: Dimensión 1, Planificación estratégica de la evaluación (C.1.1., C.1.2., C.1.3. y C.1.4.); Dimensión 2, Claridad de las condiciones y criterios de la evaluación (C.2.1.); Dimensión 4 comparable, Regulación del aprendizaje por medio de la evaluación (C.5.3., C.6.1. y C.6.2.); Dimensión 7, Participación discente en la evaluación (C.7.1.).

OBJETIVO H. Describir las experiencias y demandas que plantea el profesorado con respecto a la formación permanente vinculada a la evaluación del aprendizaje.

El último de los propósitos abordados por medio del trabajo empírico se desglosa en dos propósitos relacionados. Centrados en la capacitación evaluativa del profesorado, abordan respectivamente la formación realizada y la demandada.

OBJETIVO H1. Describir los principales rasgos de las actividades formativas realizadas por el profesorado en los últimos años con relación a la evaluación del aprendizaje.

El porcentaje de profesorado que ha llevado a cabo alguna actividad de formación permanente relacionada con la evaluación en los dos últimos cursos no representa ni la cuarta parte de la muestra (23,5%).

- La autoformación destaca como la *modalidad formativa* más extendida (46,66%), caracterizada por la iniciativa del propio sujeto para la resolución de sus inquietudes (consulta a colegas, lecturas, participación en jornadas específicas, entre otras). Parte de la formación desarrollada ha tenido lugar en módulos específicos sobre evaluación, ofertados en programas más amplios de formación de postgrado. El clásico formato “curso”, caracterizado por la presencialidad y dirección de un agente externo, y su versión en línea resultan las opciones menos frecuentes.
- La *duración* media de la formación declarada se aproxima a las 38 horas, aunque con un grado de dispersión muy elevado ($s = 34,60$).
- En el 90,9% de los casos la implicación en la formación es *voluntaria*; en el porcentaje restante se refiere una mezcla de voluntariedad y conveniencia laboral (sumar méritos, promociones, etc.), en ningún caso obligatoriedad.
- Los temas abordados en las actividades formativas giran en torno a cinco núcleos, todos ellos estrechamente relacionados con las consecuencias evaluativas derivadas del EEES e interconectados entre sí. La *evaluación de competencias* es el tema más tratado (27,77%), seguido del *planteamiento teórico general de la evaluación* y de la posible *aplicación de las tecnologías de la comunicación en la misma* (22,22%, respectivamente). Los *instrumentos* de evaluación y el enfoque de la *evaluación para el aprendizaje* completan la lista de temas trabajados.
- La valoración de la *utilidad o satisfacción* que manifiestan los docentes con respecto a la formación permanente efectuada desprende un tono positivo general. El grado de satisfacción constatado es moderado en el 44% de los casos y adecuado o muy adecuado en todos los restantes.

OBJETIVO H2. Identificar las principales necesidades formativas manifestadas por el profesorado con relación al desarrollo de su competencia evaluativa.

Los resultados hallados ponen de manifiesto que la amplia mayoría de la muestra docente (más del 81%) admite tener necesidades formativas vinculadas a la evaluación del aprendizaje discente.

- Preguntado sobre los aspectos a los que debiera orientarse dicha formación, el profesorado señala dos *núcleos temáticos* que destacan en el banco de sugerencias realizadas (el 17,85% de las aportaciones respectivamente): la evaluación de competencias y las prácticas evaluativas viables o sostenibles, relacionadas con el trabajo con grupos numerosos de estudiantes. Probablemente en conexión con ambos, *los métodos y técnicas de evaluación* representan la tercera sugerencia más reiterada (12,5%). Junto a ellas, las inquietudes vinculadas a las bases teóricas más generales de la evaluación representan otro posible foco de atención (el *proceso evaluativo*, la *profundización en los principios técnicos y/o éticos*, la *planificación evaluativa* o su *orientación al aprendizaje*). Algunas de las sugerencias minoritarias se corresponden con los factores de la competencia evaluativa analizados en el cuestionario, por lo que puede haber existido en ellas un cierto sesgo por transformación de la muestra.
- En cuanto a las *modalidades de formación preferentes*, no se constatan referencias al clásico formato curso, basado en la dirección externa y experta sobre un programa de contenidos predefinido. En su lugar, la *difusión y el intercambio de experiencias* figura como la alternativa predilecta para el profesorado, tanto en formatos presenciales como a través de redes en línea (35%). Le sigue el asesoramiento experto, pero *in situ*, es decir, orientado a la resolución de problemas y necesidades reales, identificados por las/os propias/os participantes (25%). Las experiencias de revisión o metaevaluación autoguiadas o bajo dirección externa se encuentran representadas en la misma medida entre las respuestas (25%).

CAPÍTULO 9.

COMPETENCIA EVALUATIVA DEL PROFESORADO EN LA PRÁCTICA: CONCLUSIONES DEL MARCO EMPÍRICO

Este nuevo capítulo resume las conclusiones obtenidas en respuesta a la segunda pregunta guía de la investigación, a saber, aquella que plantea identificar los rasgos que caracterizan en la práctica la competencia evaluativa del profesorado vinculado a la formación básica de maestras/os. Su exploración se ha efectuado por medio del estudio empírico realizado con 51 docentes y 704 estudiantes de los Grados de Educación Infantil o Primaria de la Universidad de Deusto y el centro adscrito BAM. Éste se ha centrado en la resolución de cuatro propósitos específicos que complementan la reflexión teórica sujeta a la primera pregunta guía de la investigación (identificar los rasgos deseables de la competencia evaluativa del profesorado).

Los próximos apartados abordan la discusión preliminar y pormenorizada de cada uno de los objetivos, contrastando los principales resultados evidenciados en su análisis y las aportaciones desprendidas de otras investigaciones y de la revisión de la literatura. El producto de la misma, junto al comunicado en el capítulo sexto para dar cierre a la indagación teórica, sustenta las conclusiones finales de la investigación, presentadas en el próximo capítulo.

1. Objetivo E. Realizar una autoevaluación de la competencia evaluativa del profesorado dedicado a la formación básica de maestras/os de Educación Infantil y/o Primaria en el contexto institucional de la UD y BAM

Conocer la valoración que hace de su competencia evaluativa el propio profesorado que la desarrolla y manifiesta representa el primer propósito del estudio empírico. Su análisis se ha concretado en una autoevaluación docente, estructurada en tres niveles de revisión complementarios: la descripción de los aspectos de la praxis evaluativa mejor y peor valorados por quienes la ejercen profesionalmente (objetivo E1), la indagación sobre los factores que condicionan tales valoraciones (objetivo E2) y la identificación de las principales fortalezas y dificultades de las que da cuenta el profesorado a la hora de evaluar el aprendizaje de las/os estudiantes (objetivo E3).

Las próximas páginas describen la discusión que suscita el contraste de las conclusiones más relevantes extraídas para cada objetivo con las reportadas en la investigación y la literatura. Su integración ofrece una visión amplia de la competencia evaluativa que declara el profesorado de la muestra, según su experiencia en los grados asociados a la formación básica de maestras/os.

1.1 Factores e indicadores de la evaluación del aprendizaje mejor y peor valorados por la muestra docente y su reflejo en la investigación y la literatura (Objetivo E1).

En congruencia con las conclusiones presentadas por Tejedor 1999 o Villa et al. (2013), en términos generales el profesorado consultado se declara competente en materia evaluativa. Sin embargo, esta primera valoración global varía sustancialmente al analizar, uno a uno, los factores e indicadores detallados en el modelo teórico propuesto, donde se reconoce un desempeño práctico insuficiente y desproporcionado a su importancia. Así sucede también en la literatura específica, que elabora un discurso en favor de la relevancia teórica de la evaluación del aprendizaje, a la vez que constata la complejidad y dificultad de su puesta en práctica a nivel técnico, ético y pedagógico (De la Orden, 2011b; Mateo, 2000; Rué, 2007, 2008).

Consideradas en su conjunto, las valoraciones recogidas de la muestra docente devuelven una imagen de la evaluación del aprendizaje a medio camino entre el estado descrito en el capítulo tercero y las orientaciones teóricas concretadas en el modelo ITC. El profesorado de las titulaciones vinculadas a la formación básica de maestras/os en BAM y la UD afirma desempeñar un uso estratégico de la evaluación, caracterizado principalmente por la planificación e implementación cuidadosa de la misma y por el empleo de sus conclusiones en la regulación del aprendizaje y la enseñanza. Aunque el grado de realización sea claramente mejorable en todos los indicadores y factores considerados (no llega a ser *muy adecuado* en ningún caso), la imagen clásica de la evaluación, como fenómeno puntual, final y acreditativo, poco fundamentado y de escaso impacto formativo, no encuentra correspondencia en el testimonio elaborado. En comparación con los criterios de calidad establecidos en el modelo ITC, las principales áreas de mejora reconocidas en la evaluación tienen que ver con cuatro focos: la participación discente, la diversificación o adecuación de condiciones en respuesta a la diversidad, la revisión de las prácticas desarrolladas y su adecuación y precisión.

Aunque los datos recopilados no permiten establecer una comparación directa, la descripción de la muestra guarda una correspondencia razonable con la tipología de profesorado ecléctico identificada por Palacios y López (2013) y Palacios et al. (2013) a partir de los rasgos del estilo evaluativo. Ésta se caracteriza por la puesta en práctica de pequeños procesos de evaluación formativa con la intervención puntual del alumnado y la combinación de procedimientos diferentes para la calificación, no sólo el clásico formato examen (Palacios et al., 2013; Palacios & López, 2013). Como presentan los propios autores de la tipología, se trata de una categoría intermedia entre el profesorado tradicional, próximo en su praxis evaluativa a las deficiencias clásicas detalladas en el capítulo tercero (evaluación final, sumativa, no participativa y centrada fundamentalmente en examen) y el innovador, que desarrolla una evaluación más acorde a la referencia de deseabilidad construida en el modelo ITC (evaluación continua y formativa, participativa, basada en una variedad de procedimientos y herramientas diversificadas y alineadas con el aprendizaje).

De modo similar, B. Gargallo et al. (2010) diferencian cuatro grupos entre el profesorado en función de las características metodológicas y evaluativas. Su clasificación parte de la clásica distinción entre el profesorado centrado en la enseñanza y el aprendizaje planteada en el primer capítulo (Barr & Tagg, 1995; García-Valcárcel, 1993; Gow & Kember, 1993; Hernández, Maquilón et al., 2010; Hernández & Maquilón, 2011; Jackson, 2006, Kember & Gow, 1994), a la que añaden dos estadios de aproximación intermedios. En el polo positivo, describen prácticas evaluativas formativas, continuas y participativas para el profesorado centrado en el aprendizaje, quien concibe la evaluación como oportunidad

didáctica. En el extremo opuesto, el profesorado centrado en la enseñanza emplea la evaluación para calificar, por lo general por medio de exámenes finales, basados en la reproducción de contenidos. Los rasgos evaluativos descritos por la muestra docente, se corresponden con las categorías intermedias, en las que se reconoce el valor formativo de la evaluación y su necesaria continuidad, aunque con limitaciones y dificultades en su puesta en práctica (ausencia de evaluación inicial, predominio de funciones acreditativas, continuidad poco sensible y mejorable, etc.).

Los resultados obtenidos en este primer objetivo (E1) dan cuenta de esa dificultad general, aunque con matices y diferencias evidentes en función del foco particular de atención considerado. De acuerdo con la exposición realizada en el capítulo anterior, los factores de la competencia evaluativa en los que menor dificultad manifiesta el profesorado son la Regulación del aprendizaje y la enseñanza, la Planificación estratégica de la evaluación y su Puesta en práctica. En el extremo opuesto, los factores vinculados a la Metaevaluación y a los Ajustes en la evaluación y la participación discente representan los más carentes según las propias valoraciones docentes.

Factores mejor representados en la competencia evaluativa docente según el profesorado...

Según el profesorado, la ***Regulación del aprendizaje y la enseñanza por medio de la evaluación*** representa el factor más destacado en importancia y el tercero más realizado, con un cumplimiento declarado levemente por encima de lo *adecuado*. Hace referencia al empleo pedagógico de la evaluación como herramienta para la mejora docente y discente, y su relevancia genera un acuerdo unánime: representa la valoración que menor grado de dispersión concentra en la muestra y es la única en la que las variables de identificación estudiadas no producen diferencias significativas (objetivo E2).

Incluida dentro del mismo, la *Regulación de la enseñanza* (ítem P.6.3.) destaca como el indicador de la competencia evaluativa más importante y realizado para la muestra docente. En línea con la utilidad y relevancia que se le reconoce desde la literatura (Boud & asociados, 2010; Brown et al., 1996; Nicol, 2007, 2009), para el profesorado representa uno de los apoyos reconocidos en la praxis evaluativa, al interpretar que la *información que se desprende de la evaluación continua del alumnado, supone una fuente de reflexión para mejorar el propio sistema evaluativo* y, por extensión, la enseñanza (objetivo E3). Así lo advirtió también Tejedor (1999) en un estudio contextualizado en la Universidad de Salamanca, donde profesorado de distintos centros y titulaciones reconoció en la evaluación

una fuente de reflexión sobre los métodos de enseñanza aplicados y el programa de la asignatura.

Junto a la regulación de la enseñanza, *la comisión de feedback* o retroalimentación (ítem C.6.1.) representa el indicador de competencia evaluativa más relevante para el profesorado, en una apreciación ampliamente suscrita en la literatura desde la década de los noventa (Black & Wiliam, 1998b; Chickering & Gamson, 1999; Hattie et al., 1996) y hasta nuestros días (Boud & Molloy, 2013; Duncan, 2007; Evans, 2013; Hattie & Timperley, 2007; Nicol et al., 2014; Nicol & Macfarlane, 2006; Orsmond et al., 2013; Yang & Carless, 2013). Desde la perspectiva teórica de la evaluación *para* el aprendizaje el feedback representa la pieza clave para contribuir al aprendizaje discente (Assessment Reform Group- ARG, 2002; Carless, 2007, 2013a; J. Chappuis et al., 2012; S. Chappuis & Chappuis, 2007; Stiggins, 2005a, 2005b; Stobart, 2010) y su presencia es unánime entre los decálogos y manifiestos que orientan y definen la praxis evaluativa de calidad⁹⁸. Esta visión parece compartida en la muestra docente estudiada, así como en la conformada por estudiantes (objetivo F1).

La transferencia de este indicador a la práctica es *adecuada* en opinión del profesorado, quien describe un feedback puntual, frecuente y suficiente, aunque no positivo ni personalizado. De hecho, la dificultad para ofrecer al alumnado información y seguimiento individualizado representa uno de los mayores obstáculos reconocidos por las/os docentes en su praxis evaluativa (objetivo E3). Los datos proporcionados desde la investigación próxima revelan una visión similar entre docentes de otras titulaciones universitarias, para quienes la retroalimentación constituye una práctica didáctica frecuente y generalizada (G. Rodríguez et al., 2012). En un estudio reciente L. F. Martínez et al. (2012) han concluido que el 68,8% del profesorado afirma ofrecer feedback siempre o bastantes veces, frecuencia a la que también se adhiere la muestra docente estudiada por Gutiérrez et al. (2013). Los resultados reportados en el contexto específico de la formación básica de maestras/os o profesorado reafirman la misma tendencia, calificando como habituales los comentarios destinados a aclarar a las/os estudiantes los resultados evidenciados en la evaluación (Arribas et al., 2010; Gutiérrez et al., 2011). Aunque el contraste de percepciones corresponde a la discusión del objetivo G, vale la pena anticipar que esta visión de la retroalimentación en la práctica no concuerda con la que pone de manifiesto el alumnado, segundo agente involucrado en su desarrollo.

⁹⁸ De acuerdo a la revisión efectuada en el capítulo cuarto, así se constata en Boud & asociados (2010), Brown et al. (1996), CLE-UB (2011), Gibbs & Simpson (2009), Grupo GRAPA-RIMA (2012), Nicol (2007, 2009), NUS (2010), Ramsden (2003), Proyecto REAP (2005) y WMG-ASKe (2007). Así lo contemplan también Cano e Ion (2012), Cano et al. (2012) y Buscà, Capllonch et al. (2012) en sus respectivas síntesis.

Naturalmente vinculada al feedback, la *sugerencia de pautas de proacción* (ítem C. 6.2.) también ocupa un lugar destacado. La literatura específica asocia este aspecto a la calidad de los procesos de los de retroalimentación y al valor pedagógico de la evaluación como recurso para el aprendizaje (Atkin et al., 2001; Hounsell, 2003; Nicol, 2013; Orsmond et al., 2013; Ramaprasad, 1983; Stiggins & Chappuis, 2005; Walker, 2013). Desde la perspectiva empírica, G. Rodríguez et al. (2012) han concluido recientemente que más del 82% del profesorado universitario consultado manifiesta ofrecer retroalimentación prospectiva o pautas de mejora, y así se evidencia también en la muestra docente analizada. Las/os integrantes de la misma lo asumen como un indicador relevante de la competencia evaluativa, de modo que declaran ofertar al alumnado pautas claras de actuación en función de los resultados evidenciados en la evaluación. Como en el caso del feedback, éstas recomendaciones son generales y se dirigen principalmente al grupo; no se producen, por tanto, las condiciones adaptativas y personalizadas que recomiendan Laurillard (2002), Jonsson (2013) o Knight y Yorke (2003) para su mayor efectividad.

La ***Planificación estratégica de la evaluación*** representa el segundo factor más relevante para el profesorado y el más llevado a cabo según sus valoraciones. Su importancia se confirma al encontrarlo referido como una de las dos principales ayudas para el ejercicio de la evaluación por las/os docentes (objetivo E3). Incluye cuatro ítems o indicadores, todos ellos calificados como *relevantes* y sujetos a un cumplimiento autodeclarado entre *adecuado* y *muy adecuado*. Cada uno de ellos recoge una de las grandes decisiones a adoptar en el diseño o planificación de la evaluación del aprendizaje: la continuidad, la coherencia o alineación, la combinación de funciones o propósitos y la previsión de instrumentos de evaluación.

La formulación de una *propuesta evaluativa alineada con los objetivos formativos y el planteamiento metodológico* (ítem C.1.2.) es el aspecto de la planificación que mejor valoración docente recibe, tanto en importancia como en realización. Los promedios de cumplimiento registrados describen prácticas muy próximas a la alineación plena, donde el límite entre las actividades de aprendizaje y evaluación es difuso, tal y como recomienda la literatura (Carless, Joughin, Liu, & asociados, 2006; Coll et al., 2007; Yániz & Villardón, 2006). Los trabajos de L. F. Martínez et al. (2012) y Gutiérrez et al. (2013) confirman esta misma autopercepción de coherencia en docentes vinculados a la formación básica de profesorado de Educación Física.

La *continuidad de la evaluación* (ítem C.1.1.) representa otra cuestión claramente importante para la muestra docente, tanto desde la perspectiva teórica, como desde la práctica. El profesorado afirma diseñar dispositivos evaluativos que incluyen tareas para la recopilación de evidencias en los tramos inicial, central y final del proceso formativo, sin

atomizar en partes la valoración del aprendizaje. Esta conclusión y su correspondencia en investigaciones similares sugieren el replanteamiento del patrón clásico de la temporalidad sujeta al proceso evaluativo. Frente al predominio de la evaluación final que reconocía buena parte del profesorado hace poco más de una década (Nieto, 2000; G. Rodríguez, 1998; Tejedor, 1999), estudios recientes recogen el compromiso práctico de las/os docentes en la articulación de sistemas de evaluación continua (Arribas et al., 2010; Olmos & Rodríguez, 2010; Villa et al., 2013). El análisis de los programas docentes y de la normativa que regula su ejercicio en la universidad reafirma esta visión, y define la evaluación como un proceso continuo, presente en las distintas fases del ciclo formativo (Ibarra & Rodríguez, 2010; Padilla et al., 2010). Con todo, advertimos ciertos indicios que llevan a pensar que pudiera estar produciéndose una interpretación restringida y limitada de dicha continuidad desde el punto de vista institucional y docente:

- En primer lugar, el promedio de cumplimiento declarado por el profesorado para la continuidad evaluativa (ítem C.1.1.) no se corresponde con el detallado específicamente para la evaluación inicial (ítem C.3.1.), que se sitúa por debajo de lo *adecuado*. Además de reafirmar una limitación ampliamente reseñada (A. Gargallo, 2009; V. M. López, 2004; Smitter & Siso, 2007), la ausencia de evaluación inicial podría indicar una interpretación limitada de la evaluación continua en términos de *evaluación no final*.
- El tipo de cuestión empleada en la investigación para indagar sobre la continuidad evaluativa también refuerza esa suposición. Desde nuestro punto de vista, los estudios sobre el tema en ocasiones parten de una interpretación limitada de la evaluación continua, que se refleja en el modo de cuestionarla ante la muestra. Así por ejemplo, en Tejedor (1999) o Clipa y Duta (2010), las preguntas planteadas para valorar la continuidad refieren sólo al número de pruebas (exámenes) efectuados a lo largo del proceso formativo. Ello reafirma la sospecha de que la evaluación continua pueda estar siendo concebida como una *práctica acreditativa ni final, ni única*.

Las valoraciones de la muestra docente para el indicador relativo a la *combinación de las funciones pedagógica y acreditativa* de la evaluación (ítem C.1.3.) atestiguan parcialmente esta aparente simplificación. Como en Arribas et al. (2010), el profesorado considera relevante la integración de ambos usos, de modo que la evaluación no sólo represente un recurso para la medición del aprendizaje y contribuya a su desarrollo. En la práctica esta combinación se percibe satisfecha de forma adecuada, aunque la evaluación no llega a constituir un instrumento para la regulación y la gestión estratégica del esfuerzo y

el trabajo discente según el propio profesorado. Se advierte un cierto predominio de la vertiente acreditativa de la evaluación, sin menoscabo de su funcionalidad formativa, y así se desprende también del análisis de los programas y normativas evaluativas (ANECA, AQU, & ACSUC, 2007b; Ibarra & Rodríguez, 2010; Jessop, 2012; Real Decreto 1791/2010).

Las anteriores consideraciones dan cuenta de la compleja relación que se produce entre las funciones de la evaluación y su continuidad en el proceso formativo. Motivado probablemente por la instalación progresiva del EEES, parece probado el aumento de la frecuencia evaluativa que ha tenido lugar en la universidad en la última década.; no así la consolidación de su misión formativa, que en la práctica sigue siendo secundaria (Villa et al., 2013). Esta circunstancia suscita en la literatura una inquietud creciente ante los posibles riesgos de incurrir en dispositivos evaluativos poco sensibles con el proceso de aprendizaje discente, que hagan de la evaluación continua una sucesión de demandas sumativas (S. Chappuis et al., 2009; Monereo, 2003; Monereo & Castelló, 2009b; Sabirón & Arraiz, 2013).

En cuanto a la *previsión de los instrumentos* a aplicar en la evaluación (ítem C.1.4.), el profesorado refiere el uso de procedimientos e instrumentos variados, capaces de proporcionar evidencias sobre el grado de desarrollo de las competencias en las condiciones prácticas previstas para su desarrollo. Aunque en el estudio no se haya incluido ninguna pregunta específica destinada a indagar sobre los procedimientos empleados, esta declaración inicial de variedad es consistente con la proliferación de herramientas evaluativas acusada en trabajos recientes (Buscà, Pintor et al., 2010; Ibarra & Rodríguez, 2010; Villa et al., 2013). No disponemos de datos para precisar si el examen sigue teniendo mayor peso y protagonismo entre las mismas, como sugieren M. L. Pérez et al. (2000) y L. F. Martínez et al. (2012) en sus respectivas investigaciones; tampoco para determinar si las demandas que plantean persisten en la evocación de contenidos más que en su aplicación, como apuntan Cid-Sabucedo et al. (2009) o Baartman et al. (2013). Los comentarios y sugerencias ofrecidas por el alumnado en el siguiente objetivo (D) ofrecen más datos en este sentido y dejan entrever la vigencia de tales apreciaciones. Según se desprende de la rúbrica de valoración aplicada, en opinión del profesorado la principal debilidad de los instrumentos dispuestos en la evaluación es su validez y fiabilidad. Esta limitación se corresponde con un área de mejora más amplia, en la medida en la que **la Adecuación y precisión** representa para la muestra docente el segundo factor de la competencia evaluativa menos garantizado en la práctica.

La **Puesta en práctica de la evaluación** cierra la lista de los factores en los que mejor se autopercebe el profesorado. Hace referencia a la *implementación coherente del*

dispositivo evaluativo diseñado, así como al desarrollo de actividades para la *difusión e interiorización de los criterios y condiciones* (ítem C.2.1.) en los que ésta se basa.

Hace más de dos décadas, J. R. Frederiksen y Collins (1989) definían la transparencia, como la necesidad de comprensión de los criterios de evaluación por parte del alumnado. Dicha orientación sigue ampliamente representada en la literatura actual (Boud & asociados, 2010; O. Esteve, 2010; Gibbs, 2006; Monereo & Castelló, 2009a) y ha pasado a ser concebida como derecho discente y obligación docente en los marcos reguladores para el ejercicio evaluativo (ANECA, AQU, & ACSUC, 2007b; ENQA, 2009; Real Decreto 1791/2010). Así se evidencia también en las escalas de evaluación de la calidad docente, donde la *claridad informativa* representa el núcleo temático con mayor representación en la muestra de instrumentos analizada (objetivo D).

El profesorado consultado admite la relevancia de este indicador de la calidad evaluativa y afirma desempeñarlo de forma solvente. En línea con las recomendaciones de J. Chappuis (2005), Carless (2006) o Price y O'Donovan (2006), no sólo detalla y explicita los criterios y condiciones que rigen el proceso de evaluación, sino que declara realizar actuaciones puntuales para favorecer su difusión y clarificación a lo largo del proceso formativo (ejemplos, recordatorios, aclaraciones, etc.). Esta apreciación pone en entredicho el clásico lamento por la falta de transparencia y la opacidad evaluativa (Darby, 2007; Dochy et al., 2002) y confirma los resultados de otras investigaciones recientes. En ellas, la presentación de los criterios de calidad y evaluación representa una práctica extendida y generalizada (Arribas et al., 2010; Gutiérrez et al., 2011; L. F. Martínez et al., 2012), siendo menos frecuente la realización expresa de actividades para su explicitación y discusión (G. Rodríguez et al., 2012). La inquietud por garantizar la transparencia evaluativa también se corrobora desde el análisis de la normativa evaluativa y de las guías y programas docentes (Ibarra & Rodríguez, 2010; Padilla et al., 2010; M. A. Sánchez et al., 2010).

Aunque la perspectiva discente ofrece un cierto contrapunto en este panorama (objetivo G), esta evolución hacia cotas mayores de claridad y transparencia responde presumiblemente a la implantación del proceso de convergencia y a la proliferación de los sistemas de garantía interna y externa de la calidad. En este contexto, vale la pena considerar que un reducido sector de la muestra docente advierte en la explicitación una traba a la flexibilidad, que obstaculiza su trabajo (objetivo E3).

En cuanto a la *puesta en práctica coherente* del dispositivo evaluativo diseñado (ítem C.4.1.), representa un aspecto importante para el profesorado consultado, aunque en un orden de menor prioridad que los factores e indicadores recorridos hasta el momento. El cumplimiento autopercebido en el mismo indica que el dispositivo evaluativo se aplica según

lo previsto y atendiendo a las principales condiciones técnicas que pueden mediar en su desarrollo. En un estudio sobre creencias pedagógicas desarrollado en la Universidad de Málaga, Serrano (2010) concluyó que los ajustes o modificaciones en el plan evaluativo se interpretan como falta o incumplimiento por buena parte del profesorado de la facultad de Educación (76,5% de la muestra). Acaso por ello, los resultados presentados en otros trabajos refuerzan la declaración de coherencia en la aplicación de lo dispuesto en el programa para la evaluación (L. F. Martínez, 2012), aunque el control de las condiciones con repercusión potencial en su efectividad técnica (el tiempo, la duración de las pruebas y tareas, etc.) no siempre representa una inquietud habitual para el profesorado (M. L. Pérez et al., 2000).

Factores peor representados en la competencia evaluativa docente según el profesorado...

Confirmando las apreciaciones recabadas de la literatura (Monereo & Castelló, 2009a; Monescillo & Toscano, 2007; Santos & Moreno, 2004), la ***Metaevaluación o revisión del sistema de evaluación*** destaca como el menos realizado de todos los factores asociados a la competencia evaluativa. También resulta el que mayor grado de dispersión concentra en ambas subescalas, lo que denota desacuerdo teórico y diferencias prácticas reseñables entre las/os docentes que integran la muestra.

La falta de revisión de los dispositivos de evaluación articulados no representa una conclusión novedosa (B. McDonald, 2010; C. L. Rodríguez, 2008). Aun así, resulta llamativa la desproporción entre el escaso cumplimiento declarado por el profesorado para este factor y el grado relevancia que se le estima. Más aún si se considera que la Metaevaluación subyace a varias de las ayudas señaladas por las/os docentes en el ejercicio de la evaluación del aprendizaje (revisión anual de los programas, valoraciones del alumnado sobre los dispositivos evaluativos configurados) (objetivo E3).

En la propuesta de sistematización presentada la Metaevaluación incluye la *revisión del dispositivo evaluativo configurado desde la perspectiva ética* (ítem P.8.3.), *técnica* (ítem P.8.2.) y *pedagógica* (ítem P.8.1.). Aunque no de manera regular ni sistemática, la metaevaluación de la dimensión ética (justicia, privacidad, transparencia y beneficencia de la evaluación) destaca como la más importante y llevada a cabo, según el profesorado. No así la supervisión de los principios de viabilidad, sistematicidad, validez y fiabilidad vinculados a la revisión técnica, donde se produce el menor de los promedios de cumplimiento declarados. Este dato respalda nuevamente la valoración que realiza la

muestra del factor **Adecuación y precisión**, el siguiente menos llevado a cabo después de la metaevaluación. La reflexión sobre la consecuencialidad de la evaluación o su utilidad para contribuir al aprendizaje (Gielen et al., 2003; Hargreaves, 2007; Monereo & Castelló, 2009a, 2009b; Rotger, 1992), es la menos importante de las tres para el colectivo docente. De acuerdo con Stiggins (1998), esta falta de revisión y reflexión pedagógica supone mayores riesgos y repercusiones en las instituciones dedicadas a la formación del profesorado, y así lo hemos manifestado también en otros foros (Bilbao, 2010, 2011b).

En la lista de los factores peor representados en la competencia evaluativa autopercibida por el profesorado, el **Ajuste de la evaluación y la participación discente** destaca por ser el menos valorado en relevancia. La estructura interna revelada en las respuestas de la muestra incluyen en él cuatro indicadores referidos a dos núcleos temáticos diferentes: la participación del alumnado en la evaluación (ítem C.7.1.) y la atención a la diversidad (ítems C.3.1., C.3.2. y P.4.2.).

Los resultados puestos de manifiesto con relación al segundo núcleo sitúan la atención a la diversidad como un espacio de grandes controversias en el terreno específico de la evaluación del aprendizaje. De acuerdo con V. Álvarez et al. (2002) la personalización representa un reto de mejora general en la enseñanza universitaria. Las conclusiones extraídas del presente análisis sugieren que ese reto resulta especialmente evidente al abordar la evaluación, profundamente apegada a la homogeneización y a la uniformidad. De hecho, la *atención individualizada* representa para el profesorado el segundo obstáculo más importante en el desempeño de su competencia evaluativa (objetivo E3). Incluso en aquellos espacios en los que la muestra se considera más solvente (comisión de feedback y pautas de proacción) la falta de personalización se admite de forma expresa (comentarios generales, pautas de mejora globales, etc.).

La *adecuación de la evaluación en respuesta a la diversidad del alumnado* (ítem C.3.2.) constituye para las/os profesoras/es el indicador menos relevante de la competencia evaluativa y el que mayor grado de dispersión o desacuerdo concentra. Esta visión de la diversificación de los sistemas evaluativos como algo secundario y conflictivo se ve reforzada en indicios procedentes de otras investigaciones. M. L. Pérez et al. (2000) y Tejedor (1999), por ejemplo, mostraron en sus trabajos que el personal docente vive como preocupación la adecuación y flexibilización de los criterios y condiciones evaluativas. Incluso en la consulta a expertas/os desarrollada en el transcurso de la presente investigación, la diversificación de la evaluación ha sido destacada como uno de los componentes menos prioritarios y más controvertidos de la competencia evaluativa.

Lejos de la visión compensatoria que sostienen Verdejo (2008) o Baker et al. (1996), la justicia en la evaluación parece interpretarse en términos de objetividad e igualdad. De hecho, la no discrecionalidad de condiciones y premisas evaluativas se incluye en el repertorio de rasgos deseables en la docencia universitaria (V. Álvarez et al. 1999); así se constata en los tópicos considerados en las escalas institucionales para su evaluación, donde la justicia aparece asociada al rigor y la precisión (objetivo D). Esta misma visión también se refleja en la normativa evaluativa, donde la estereotipia se sitúa por delante de la adecuación, reservada a situaciones de excepcionalidad (Padilla et al., 2010). Mientras que la igualdad y publicidad de la calificación del aprendizaje aparece claramente contemplada como deber docente y derecho del alumnado (ENQA, 2009; Real Decreto 1791/2010), el establecimiento de medidas compensatorias no está reflejado salvo para regular circunstancias extraordinarias en la realización de exámenes (Real Decreto 1791/2010).

En correspondencia con tales apreciaciones, el profesorado de la muestra describe un ejercicio puntual y extraordinario de flexibilización, lejos de la aplicación de itinerarios evaluativos alternativos que ejemplifican Navarro (2010) o V. M. López (2008) (este último en el contexto específico de la formación básica de maestras/os). Interpretados en suma, estos resultados ofrecen un contrapunto práctico al discurso teórico que define la diversificación como indicador de buena praxis evaluativa en la universidad (Cano et al., 2010; Chickering & Gamson, 1999; Mateo, 2006b, 2008) y en etapas educativas inferiores (Chickering & Gamson, 1987; Sanmartí, 2007; Santos, 2006).

El indicador referido a la *participación del alumnado en la evaluación* (ítem C.7.1.) representa a juicio docente el segundo menos llevado a cabo. La muestra consultada refiere prácticas participativas puntuales, asociadas a la valoración de trabajos propios o de compañeras/os, que no afectan a la planificación de la evaluación ni a su revisión. Igual que en el estudio de C. Brew et al. (2009), una pequeña fracción del profesorado reconoce los ejercicios de coevaluación o evaluación entre iguales como una de las ayudas en su praxis evaluativa (objetivo E3), en la medida en la que alivian la carga asociada a la corrección.

A pesar de los muchos testimonios que respaldan y avalan su valor pedagógico (Dochy et al., 1999; Falchikov & Boud, 1989; V. M. López, 2008; V. M. López, Fernández-Balboa et al., 2012; Ureña et al., 2009; Vu & Dall'Alba, 2007; Watts & García-Carbonell, 2006), la evaluación participativa no parece extendida en la muestra consultada ni en otras similares. Alegando recelos o desconfianza ante la capacidad del alumnado para generar valoraciones basadas en criterios externos o estándares profesionales específicos (G. Rodríguez et al., 2012; Zaragoza et al., 2009), el profesorado reconoce las prácticas de

autoevaluación, coevaluación o evaluación compartida como puntuales y anecdóticas. Los programas docentes y los marcos reguladores para la evaluación en la universidad reafirman esta visión hegemónica de la evaluación, donde la intervención discente no aparece contemplada en ninguna de las fases del proceso (Ibarra & Rodríguez, 2010; Padilla et al., 2010; G. Rodríguez et al., 2012). Confirmando las apreciaciones de Trillo y Porto (2002), las escalas de evaluación de la calidad docente tampoco reflejan la intervención discente como un núcleo de atención prioritaria (objetivo D), de lo que se deduce su escasa valoración como factor de calidad docente y evaluativa.

Reservar la evaluación a la intervención docente resta a la misma potencialidad didáctica, muy en especial en los grados vinculados a la formación inicial de profesorado, comprometidos en la capacitación evaluativa (I. Álvarez, 2008a; V. M. Castejón et al., 2012; V. M. López et al., 2007). Sin embargo, parece representar una rasgo ampliamente extendido y generalizado; así lo manifiesta el acuerdo que al respecto manifiestan estudiantes y profesorado (objetivo G), tanto en la preparación de maestras/os o profesorado (Arribas et. al, 2010; Gutiérrez et al., 2011; Gutiérrez et. al., 2013; Ruiz-Gallardo et al., 2013) como en otras titulaciones (Porto, 2006). Desde esta perspectiva, resultan más esperanzadoras las conclusiones de C. Brew et al. (2009), quienes reportan un incremento de prácticas evaluativas participativas en la formación básica de maestras/os.

1.2 Relación entre variables socio-demográficas, académicas y profesionales y la competencia evaluativa autopercebida: conclusiones y reflejo de las mismas en la investigación y la literatura (Objetivo E2).

La imagen general descrita para la competencia evaluativa docente en las valoraciones de relevancia y cumplimiento autopercebido varía en función de determinadas variables. Los resultados puestos de manifiesto en respuesta al objetivo E2 revelan un amplio abanico de relaciones e influencias. Detalladas a lo largo del capítulo octavo, resultan de especial relevancia en este apartado aquellas referidas a la formación previa y permanente del profesorado. Su incidencia reiterada en las diferencias y asociaciones significativas halladas y la magnitud del efecto asociado a las mismas, así lo justifican. Frente a ellas, la formación docente de postgrado se descarta en la muestra como variable

implicada en la calidad de la praxis evaluativa, respaldando una conclusión formulada por Palacios y López (2013) para el mismo contexto formativo.

En contra de lo manifestado por Caballero (2013) y Palacios y López (2013), la **formación previa** del profesorado representa el tipo de preparación que más y mayores diferencias suscita al comparar sus valoraciones. Las discrepancias más reseñables se concentran en dos de los grupos profesionales, que muestran una tendencia desigual de respuesta. Por un lado, las/os tituladas/os en *Pedagogía, Psicología y Psicopedagogía* tienden a puntuar significativamente por encima que el resto del profesorado la importancia y la realización atribuida a varios de los factores asociados a su competencia evaluativa. Las diferencias registradas resultan estadísticamente relevantes y de tamaño grande en el cumplimiento total autodeclarado y en el referido a dos factores en particular: la Planificación estratégica de la evaluación y los Ajustes de la misma y la participación discente. Por otro lado, las/os tituladas/os en *Filología* detallan un nivel de cumplimiento significativamente menor y de efecto grande con respecto al resto de docentes en las mismas valoraciones. Los límites reconocidos para la investigación en curso no permiten delimitar si estos resultados se deben realmente a un mejor desempeño o a un grado desigual de autoexigencia en la valoración. Advertimos en este último caso que, por su mayor especialización didáctica, el profesorado con formación previa psicopedagógica o pedagógica pudiera estar sujeto a una mayor presión por la deseabilidad social que el titulado en Filología.

Las conclusiones puestas de manifiesto en investigaciones similares no ofrecen respaldo unánime a las diferencias reportadas. Existen precedentes en los que se señala la configuración de diferentes culturas de trabajo en función del área de conocimiento (V. Álvarez et al., 2002; Gil-Flores, 2012). Aunque los datos no permitan una equiparación directa, Tejedor (1999) evidenció prácticas evaluativas significativamente diferentes en titulaciones vinculadas a distintas áreas de conocimiento, donde presumimos una formación previa dispar entre el profesorado. Estudios comparativos recientes detectan variaciones en el tipo de demandas de evaluación, en la cantidad de retroalimentación ofrecida o en el tipo de herramientas empleadas y su ponderación en función de la disciplina de referencia (Jessop & Maleckar, 2014; N. Martínez et al., 2012). Sin embargo, Porto (2006) concluyó que la praxis evaluativa no difiere en las distintas titulaciones, lo que contravendría la influencia potencial de la formación previa docente. En la misma dirección apuntan las contribuciones de Serrano (2010), que mostró que las creencias pedagógicas sobre evaluación del profesorado de la Facultad de Educación de la Universidad de Málaga no varían en función de su formación previa.

Ante la falta de consenso empírico, la discusión teórica sobre el papel de la formación previa en la práctica evaluativa también ofrece aportaciones relevantes a este debate. La literatura da cuenta de numerosos testimonios que señalan las repercusiones negativas de la falta de formación psicopedagógica específica en el acceso a la docencia universitaria (V. Álvarez et al., 2011; Caballero, 2013; Delbanco, 2012; A. Fernández, 2008; Margalef & Álvarez, 2005; Mateo, 2012; Monereo, 2013, 2014a; Torra et. al., 2013; Zeichner, 2010). La tesis que subyace a tal postura es el reconocimiento de la enseñanza como proceso complejo, que no es abarcable sólo con recursos y saberes de carácter disciplinar (J. M. Esteve, 2006; Montero, 2006; Zabalza, 2002). En la propia muestra analizada *la falta de preparación específica* figura entre los aspectos que más dificultan la praxis evaluativa (objetivo E3), atribución también presente en trabajos previos (Flórez, 1999; Nieto, 2000). La declaración de mayores niveles de realización en aquellas/os profesionales con formación previa en el área Pedagógica, Psicológica o Psicopedagógica, donde se presume una mayor capacitación psico-didáctica, representa un apoyo indirecto en favor de dicha tesis. Así lo advertimos también en las observaciones de Aparicio y Pozo (2006), quienes sostienen que el profesorado con mayor formación pedagógica manifiesta concepciones más ricas del aprendizaje (por extensión, deducimos que también sobre su evaluación).

Además de aparecer referida como una de las ayudas con las que cuenta el profesorado para el ejercicio de su competencia evaluativa (objetivo E3), la **formación permanente sobre evaluación** también aparece asociada a diferencias estadísticamente significativas en las valoraciones registradas. Con un efecto medio, de menor magnitud que el constatado para la formación previa, éstas se manifiestan en tres valoraciones: el cumplimiento total autopercebido y la importancia y la realización atribuidas al factor centrado en los Ajustes y la participación discente en la evaluación. En todas ellas, las puntuaciones inferiores se registran en el grupo docente que no ha realizado ninguna actividad formativa sobre evaluación en los dos últimos años.

A diferencia de las conclusiones expuestas, Palacios y López (2013) sugieren que la formación permanente tiene más repercusión que la titulación inicial del profesorado en el modo de llevar a cabo la evaluación. Así lo afirman al constatar que el profesorado con titulación psicopedagógica tiene una representación mayor en los grupos de prácticas evaluativas tradicionales y eclécticas, mientras que el perfil innovador se compone principalmente de docentes tituladas/os en otras especialidades, que han llevado a cabo actividades de formación permanente en cuestiones didáctico-pedagógicas.

De un modo u otro, con menor o mayor prioridad, la conclusión resultante es que la formación docente, inicial y permanente, representa un factor clave en la consolidación de la

competencia evaluativa. A menos que se modifiquen las condiciones que regulan el acceso profesional a la enseñanza universitaria, el foco prioritario de actuación se concentra en la formación permanente. Entendida como refuerzo o complemento a la formación profesional previa, ésta debería adoptar una orientación psicopedagógica según los resultados presentados, más centrada en la didáctica que en la propia disciplina. La discusión del último de los propósitos (objetivo H) ofrece una descripción pormenorizada de las temáticas y las modalidades de trabajo que sugiere para la misma el profesorado.

Las diferencias y asociaciones halladas a partir de las variables socio-demográficas, académicas y profesionales restantes aparecen sujetas a una menor repercusión dentro de la muestra, y apenas encuentran reflejo en la investigación próxima. Excluidos los resultados asociados a la formación, entre ellas sólo se detecta una única diferencia con efecto de magnitud grande. Ésta se produce a partir de los años de ejercicio en la enseñanza universitaria y se manifiesta en la valoración del grado de Regulación del aprendizaje y la enseñanza por medio de la evaluación; se trata de una discrepancia parcial, ya que sólo afecta a los dos grupos de mayor experiencia en la muestra, sin implicación del de menor recorrido docente. Confirmando las apreciaciones de Toledo (2006), Buendía y López (1999) o Serrano (2010), la experiencia no da lugar a ninguna otra diferencia o asociación significativa.

1.3 Obstáculos y apoyos en el ejercicio de la evaluación según la muestra docente y su reflejo en la investigación y la literatura (Objetivo E3).

La discusión en torno a la competencia evaluativa autopercebida por el profesorado no puede concluir sin hacer referencia a las condiciones reales en las que se desempeña. La evaluación del aprendizaje discente se lleva a cabo en un contexto determinado en el que inciden factores de naturaleza diversa que pueden obstaculizar o favorecer su desarrollo. El punto de partida de esta reflexión es que el sector docente identifica más dificultades que apoyos en el ejercicio de sus funciones evaluativas. Las principales **ayudas** señaladas son la *Planificación de los dispositivos evaluativos* y la *formación en evaluación*, ambas referidas en puntos anteriores de la discusión. Entre las dificultades se incluyen factores que condicionan las decisiones del profesorado en este terreno, y que ya han sido señalados como causa de inadecuación de los sistemas evaluativos en estudios previos (L. F Martínez et al., 2012; Porto, 2006; Tejedor, 1999).

Al revisar los principales **obstáculos** que entorpecen la práctica evaluativa, las respuestas recopiladas evidencian un amplio consenso dentro y fuera de la muestra. La *sobrecarga*, categoría en la que se incluyen toda una serie de factores con incidencia en el volumen y la temporalidad de la tarea docente (número de estudiantes, sobrecarga de funciones, carga de trabajo asociada a los sistemas de evaluación formativa y continua, condiciones prácticas de la evaluación y cambio en el encargo docente de un curso a otro), representa la dificultad más extendida y reconocida por el profesorado. Suscribiendo una apreciación ampliamente referida en trabajos anteriores, el número de estudiantes destaca como el obstáculo más señalado (Cano & Ion, 2012; Gutiérrez et al., 2013; M. L. Pérez et al., 2000; Tejedor, 1999; Zaragoza et al., 2009). La falta de tiempo y el incremento del volumen de trabajo que ocasionan los dispositivos de evaluación formativa y continua también representan una limitación evidente para las/os docentes. Así lo confirman otros estudios recientes, en especial, desde que arrancara la implantación del EEES (Capllonch et al., 2009; Navarro et al., 2010; Zaragoza et al., 2009). En el contexto internacional también se reconoce la importancia de estas dificultades prácticas, señalando la previsión y el control de su incidencia como un factor clave en el diseño de evaluaciones efectivas (Assessment Resource at Hong Kong University- AR@HKU, 2009d; Suskie, 2006).

Junto a la sobrecarga, la muestra analizada señala otras fuentes de dificultad que ya han sido mencionadas en el transcurso de la discusión anterior. Entre ellas se encuentra la *atención individualizada*, en la medida que complica operativamente la evaluación desde la perspectiva de la sobrecarga (volumen de trabajo, cantidad de estudiantes, etc.). A ella se une la también reseñada *falta de formación*, que a juicio de las/os participantes representa otro obstáculo que condiciona prácticas evaluativas inadecuadas. Como en L. F. Martínez, et al. (2012), la *complejidad natural asociada a la evaluación de competencias* figura como una de las causas que motiva especialmente dicha dificultad.

Las respuestas recopiladas de la muestra incluyen otras dificultades referidas en estudios próximos, aunque con menor representación que las tres anteriores. Tal es el caso de la falta de *coordinación docente en materia evaluativa*, también recogida como obstáculo por Porto (2006, 2010), Cano e Ion (2012) o Arribas et al. (2010). También de la *asistencia irregular del alumnado* y su incondicional derecho de evaluación, que suscita dificultades para consolidar la generalización de los sistemas evaluativos integrados en la dinámica cotidiana del aula (Capllonch et al., 2009; Navarro et al., 2010).

2. Objetivo F. Realizar un diagnóstico de la competencia evaluativa del profesorado dedicado a la formación básica de maestras/os de Educación Infantil y/o Primaria en la UD y BAM desde la perspectiva de las/os estudiantes

El segundo objetivo abordado por medio del marco empírico de la investigación se ha centrado en la percepción que manifiesta el alumnado sobre la competencia evaluativa del profesorado. La pertinencia de su intervención como parte informante en procesos de evaluación de la calidad docente y de la formación ha sido ampliamente reforzada en diferentes trabajos de revisión de la literatura (Aleamoni, 1999; Brodie, 1998; Cisneros-Cohernour & Stake, 2010; Marsh, 1984, 1987; Richardson, 2005; Williams & Brennan, 2003). A pesar de los sesgos de imprecisión de los que también dejan constancia dichos trabajos, la finalidad estrictamente formativa de la presente investigación hace del alumnado una fuente privilegiada de consulta. Su visión puede ser contrastada con la procedente de la muestra docente, en la que también cabe esperar un cierto sesgo por autocomplacencia o deseabilidad social en la autoevaluación.

El análisis de esta segunda perspectiva sobre la evaluación del aprendizaje ha sido ordenado en torno a tres propósitos específicos, igual que en la autoevaluación docente. Juntos contemplan la descripción de los aspectos de la praxis evaluativa mejor y peor valorados (F1), la identificación de los factores con mayor relación o influencia en dichas apreciaciones (F2) y la enumeración de las principales fortalezas y áreas de mejora de las que dan cuenta las/os estudiantes en la evaluación de su aprendizaje (F3).

2.1 Factores e indicadores de la evaluación del aprendizaje mejor y peor valorados en la muestra discente y su reflejo en la investigación y la literatura (Objetivo F1).

La evaluación representa una de las inquietudes prioritarias para el alumnado universitario, destacando como el tema que mayor número de comentarios suscita en el estudio de la calidad docente (Rué, 2009b). Sin embargo, esta especial sensibilidad parece responder más a la gravedad de las implicaciones y repercusiones prácticas que entraña, que a su valor formativo. De hecho, numerosos estudios revelan sensaciones negativas

con respecto a la evaluación entre el alumnado (L. F. Martínez et al., 2012; Ricoy & Fernández-Rodríguez, 2013; Ruiz-Gallardo et al., 2013), hasta el punto de situarla como el aspecto menos satisfactorio de la formación universitaria (Buckley, 2012; I. González, 2003; Higher Education Founcil Council for England- HEFCE, 2014).

Aunque su especial sensibilidad pedagógica pudiera hacer prever una situación diferente, esta imagen de la evaluación también se extrapola al contexto específico de la formación básica de maestras/os. Basados en el testimonio de alumnado actual y egresado, Hamodi y López (2012) pusieron de manifiesto una ruptura evidente entre la teoría y la práctica evaluativa en el mismo. Las valoraciones procedentes de la muestra discente apuntan en esa misma dirección, señalando diferencias evidentes entre la importancia de los indicadores asociados a la competencia evaluativa del profesorado y su grado de realización. Este desequilibrio no se manifiesta por igual en todos los factores considerados en el análisis. Las mayores dificultades tienen lugar en la Participación discente y en la Regulación del aprendizaje por medio de la evaluación, factores en los que el promedio de cumplimiento estimado por el alumnado no alcanza el nivel de lo *adecuado* en el modelo de referencia. Este requisito se satisface en los cuatro factores restantes, especialmente en el relativo a los Ajustes en la evaluación y en el Control de las condiciones durante su desarrollo.

Factores peor representados en la competencia evaluativa docente según el alumnado...

Las/os estudiantes de la muestra consideran que su **Participación en las distintas fases del proceso evaluativo** representa el factor menos relevante de la competencia docente y también el menos llevado a cabo. En el marco teórico de la investigación, la evaluación participativa ha sido presentada como uno de los ejes en los que reposa el enfoque teórico de la evaluación para el aprendizaje (ARG, 2002; Carless, 2007; J. Chappuis et al., 2012; Stiggins, 2005, 2005b) y como requisito para la consolidación de la evaluación basada en competencias (Bretones, 2008; McDowell & Sambell, 2003; Sayós, 2010). Sin embargo, su incidencia real en la práctica por el momento no parece respaldada en esta investigación, ni en otras similares (Gil-Flores, 2012).

La descripción empírica de la participación y el reparto del poder en el proceso evaluativo dista mucho de la recomendación en favor de una evaluación más autónoma y democrática formulada desde la literatura (Boud & Falchikov, 2006; Dochy et al., 1999; Falchikov & Goldfinch, 2000; A. Pérez et al., 2009). Confirmando las conclusiones de

estudios previos desarrollados en el contexto específico de la formación básica de maestras/os o profesorado (Arribas et al., 2010; Gutiérrez et al., 2011, 2013; V. M. López & Palacios, 2012; L. F. Martínez et al., 2012; Ruiz-Gallardo et al., 2013; Smitter, 2006), el alumnado declara *posibilidades de intervención* puntuales, centradas en la corrección de trabajos propios o ajenos durante la fase de implementación (C.7.1.). Advierte que el diseño de la evaluación corre a cargo del profesorado y que la fase de *metaevaluación ética, técnica y pedagógica* (E.8.1., E.8.2., E.8.3.) transcurre sin requerir su opinión. De reciente publicación, el estudio de G. Rodríguez et al. (2012) sobre la evaluación participativa en distintas titulaciones universitarias ofrece una imagen similar, caracterizada por la implicación puntual del alumnado en labores de corrección durante el proceso formativo. En correspondencia a ello, la información publicada en programas y guías didácticas devuelve una visión tradicional de la evaluación, focalizada en la figura docente (Padilla et. al, 2010). Según Ibarra y Rodríguez (2010) así se constata también al analizar el discurso dominante en la normativa evaluativa.

Además de responder a las exigencias del actual modelo universitario, la ampliación del margen de intervención en el terreno de la evaluación responde a una demanda recurrente por parte del alumnado (A. Brew, 2003; Ureña et al., 2009). Las/os estudiantes manifiestan una actitud favorable para ejercer mayor protagonismo en el proceso (Porto, 2006; G. Rodríguez et al., 2012; Tejedor, 1999), en especial en las titulaciones asociadas a la formación del profesorado, donde los beneficios didácticos y pedagógicos les resultan especialmente evidentes (V. M. López et al., 2007; McGarr & Clifford, 2013). Sin embargo, esta predisposición positiva no se refleja plenamente en los resultados obtenidos en la presente investigación, donde el colectivo discente muestra un interés moderado por las fórmulas participativas de evaluación. Señalado como el menos importante de los indicadores asociados a la competencia evaluativa, su proliferación no representa para la muestra una demanda de cambio prioritaria. Así se deduce de su escasa presencia entre las sugerencias de mejora presentadas (objetivo F3).

El cuestionario empleado no profundiza en las causas de esa aparente indiferencia, pero la investigación específica ofrece indicios relevantes para apoyar su interpretación. McMahon (2010), por ejemplo, evidenció ciertas resistencias ante la participación en el proceso evaluativo, particularmente cuando ésta tiene lugar entre iguales. A menudo se considera injusta y poco profesional (Wilson et al., 2015). El alumnado no reconoce la valoración del trabajo ajeno dentro de su cometido y tampoco se siente capacitado para desempeñarla, lo que le genera incomodidad e intranquilidad en ambos roles, en el de agente evaluador y en el de evaluado (Boud et al., 1999; Hamodi & López, 2012). Incluso admitidos los beneficios de la evaluación participativa (C. Brew et al., 2009; Hamodi &

López, 2012), este tipo de situaciones resultan impopulares y estresantes para las/os estudiantes (Falchikov & Boud, 1989; Pope, 2005); las fórmulas de coevaluación o co-calificación pueden representar una fuente de conflictos, llegando a ocasionar fricciones y tensiones en el grupo (Heathfield, 2003; Vu & Dall'Alba, 2007). Los resultados evidenciados en la presente investigación, podrían haberse originado desde una visión similar, consciente de los beneficios potenciales (hasta el punto de calificar la participación como *relevante*) y también de sus riesgos (situándola a la cola en cuanto a prioridad).

El segundo núcleo problemático que señala la muestra discente en el análisis de la evaluación es la **Regulación del aprendizaje**, centrado en la *comunicación de los resultados de la evaluación* (ítem C.5.3.) y en los procesos de *feedback* y *proacción* emprendidos a partir de la misma (ítems C.6.1. y C.6.2.). Su definición como área de mejora se justifica por el grave desajuste que concentra, con una atribución de relevancia teórica especialmente reseñable y una insuficiente manifestación práctica.

Igual que en la muestra docente (objetivo E1), el alumnado respalda el protagonismo que concede la investigación y la literatura al proceso de retroalimentación y lo destaca como el componente más importante de la competencia evaluativa. Sin embargo, las valoraciones sobre su cumplimiento denotan una manifestación práctica insuficiente o *poco adecuada*, lo que genera insatisfacción entre las/os estudiantes. Así se evidencia en la propuesta final de áreas de mejora para la praxis evaluativa, donde el *feedback deficiente* representa el tema principal en las observaciones recopiladas de este sector de la muestra (objetivo F3).

De acuerdo con lo establecido en el marco teórico de la investigación, el proceso de regulación a partir de los resultados de la evaluación se desarrolla en dos niveles de comunicación interdependientes y complementarios: el *feedback o retroinformación*, que debe ayudar al alumnado a comprender su situación actual con respecto a los logros previstos, y la *proacción*, caracterizada por la sugerencia de propuestas o vías de actuación para avanzar en dirección a los mismos. Aunque no siempre sujeta a una misma terminología, la combinación de ambos niveles de regulación está presente en modelos explicativos del efecto formativo de la evaluación desde Ramaprasad (1983) y hasta la nuestros días (Atkin et al., 2001; Hattie & Timperley, 2007; Hounsell, 2003, 2007; Sadler, 2010; Orsmond et al., 2013). Su interdependencia provoca que la proacción no pueda tener lugar con garantías de éxito, a menos que se satisfaga razonablemente la comunicación directa de los resultados y el feedback que favorece su comprensión.

El testimonio recabado de la muestra discente sugiere dificultades en el acceso al primer nivel de comunicación, en cuanto a la comprensión de las valoraciones que se

realizan de su aprendizaje. Tanto en las cuestiones cerradas, como al exponer su opinión de forma abierta (objetivo E3), el alumnado afirma que la comunicación de resultados se lleva a cabo con garantías de confidencialidad, pero en insuficiente profundidad y detalle. Igual que en otros estudios (Gil-Flores, 2012; Tejedor, 1999), se constata que, por lo general, la información proporcionada es numérica. Según las observaciones recopiladas, esta comunicación resulta poco puntual e infrecuente, de forma que la *ausencia de resultados parciales sobre la evolución del aprendizaje* representa la segunda debilidad más reiterada (objetivo F3). Cuando las cuantificaciones se acompañan de comentarios, las/os estudiantes advierten dos dificultades ligadas a su grado de generalidad, ambas reportadas por Ferguson (2011) en un trabajo previo:

- La primera de ellas apunta a la amplitud e imprecisión de los mensajes recibidos que, a su juicio, no permite comprender realmente los aspectos clave en los que se precisa mayor atención. Siguiendo las orientaciones que se desprenden de la literatura, parte de esa dificultad podría deberse a limitaciones en la decodificación de los estándares de calidad y las exigencias disciplinares (Carless, 2006; Higgins et al., 2002; National Union Students- NUS, 2010; Price et al., 2010; Sadler, 2013); sin embargo, las aportaciones de la muestra apuntan más bien a un insuficiente nivel de desarrollo (objetivo F3).
- Siguiendo la tendencia general que señalan para la formación universitaria V. Álvarez et al. (2002), la segunda dificultad hace referencia a la falta de personalización del feedback, que se percibe desligado de la particularidad y más orientado al grupo que al individuo. Frente a esta constatación, la investigación específica revela una clara preferencia del alumnado por los formatos personalizados (Huxham, 2007; Jonsson, 2013; NUS, 2010; Poulos & Mahony, 2008).

Por extensión, las pautas y sugerencias de mejora asociadas a la proacción, manifiestan los mismos síntomas según la muestra. Se describen como infrecuentes e insuficientemente generalizadas, pero cuando se producen, el alumnado las valora y aprecia; de hecho la *posibilidad de solicitar aclaración sobre las tareas o exámenes realizados* y la de *introducir correcciones y modificaciones* en los trabajos a partir de las orientaciones recibidas figuran entre los mantener en la evaluación (objetivo F3).

La imagen descrita por las/os participantes para el feedback y la proacción se sitúa a gran distancia de los criterios e indicadores de calidad establecidos en el modelo ITC. A la luz de las conclusiones publicadas en investigaciones similares, tal constatación no responde a una dificultad puntual y aislada dentro de la muestra. Jessop (2012) y G.

Rodríguez et al. (2012), por ejemplo, concluyen que en opinión discente la retroalimentación en la universidad llega tarde, limitando poderosamente las oportunidades de proacción prospectiva. En las titulaciones vinculadas a la formación básica de maestras/os o profesorado el alumnado también manifiesta una insatisfacción similar, al considerar que recibe insuficiente información sobre sus logros y avances y las posibilidades de mejorarlos (Arribas et al., 2010; Gutiérrez et al., 2011, 2013; L. F. Martínez et al., 2012). Partiendo de esta situación, la Unión Nacional de Estudiantes del Reino Unido, demanda la integración auténtica del feedback en el proceso formativo por medio de un decálogo. En él incorpora toda una serie de sugerencias para fortalecer la presencia real de la retroalimentación en la actividad formativa (feedback frecuente, puntual, vinculado a distintos tipos de evaluación) y para aproximarla a los formatos preferentes para su generación y administración (feedback variado, personal y directo al comienzo del proceso, y electrónico en el transcurso, respetuoso con el derecho a anonimato, etc.) (NUS, 2010).

Por todo lo anterior, entendemos que la reducción de la distancia entre la práctica real del feedback y la proacción y el referente de calidad establecido en el modelo ITC debería ocupar un lugar prioritario entre las líneas de mejora a emprender en materia evaluativa. Su consecución debería afrontarse progresivamente, resolviendo en primer lugar las carencias más elementales asociadas a la claridad, suficiencia, frecuencia y prospección del feedback, y atendiendo posteriormente a retos más exigentes. Entre estos últimos se encuentran aquellos indicadores de calidad incluidos en la teoría evaluativa de referencia (modelo ITC), sobre los que la presente investigación no ha profundizado específicamente. Según se desprende de la literatura, los más relevantes parecen la superación de la clásica negatividad contenida en la retroalimentación (Gibbs & Simpson, 2009; Huxham, 2007; Wootton, 2002) o la proliferación de formatos de feedback dialógico, que contribuyan a incrementar su potencialidad pedagógica y, al mismo tiempo, su sostenibilidad (Cano, 2014; Carless, 2013b; Carless et al., 2011; Sadler, 2010; Sambell, 2011). Determinadas evidencias recabadas en la muestra docente apoyan la necesidad de mejorar esa sostenibilidad, entendida en términos de viabilidad práctica. Las más claras son el reconocimiento expreso de dificultades asociadas a la *sobrecarga* y su efecto en la *comisión de retroalimentación personalizada* (objetivo E3) y la demanda directa de formación en fórmulas de *feedback sostenible* (objetivo H3).

Factores mejor representados en la competencia evaluativa docente según el alumnado...

Frente a la participación y la regulación del aprendizaje, las/os estudiantes consideran que los **Ajustes en la evaluación** es el factor mejor representado en la competencia evaluativa del profesorado. Como en L. F. Martínez et al. (2012), las respuestas registradas indican que los *compromisos adoptados con respecto al sistema de evaluación* se cumplen en la práctica casi siempre, sin que se constaten alteraciones relevantes sobre las condiciones definidas (ítem E.4.1.1.). Desde el punto de vista del alumnado, el profesorado *justifica las pocas modificaciones* que tienen lugar (ítem E.4.1.2.), siendo el *consenso de las mismas* el aspecto que en menor medida se considera cubierto (ítem E.4.1.3.). En línea con lo manifestado por la muestra docente (objetivo E), las/os estudiantes valoran la importancia del cumplimiento de las previsiones publicadas y parecen percibir las modificaciones como la excepción a la norma. En consecuencia, no disponemos de datos para reafirmar las observaciones de Smirer (2006) o el CSHE (Centre for the Study of Higher Education, 2002) sobre la demanda discente de una mayor flexibilidad. La *rigidez de los dispositivos evaluativos* aparece recogida entre las áreas de mejora atribuidas a la evaluación (objetivo F3), pero su presencia resulta minoritaria.

El **control de las condiciones de la evaluación** también representa para la muestra un factor relevante y cubierto razonablemente. Relacionado con la puesta en práctica de la evaluación, hace referencia a la supervisión que lleva a cabo el profesorado de las *condiciones prácticas y ambientales* en su desarrollo (ítem E.4.1.2. y E.4.1.3.), así como de la *legitimidad* (ítem E.4.1.1.). En contra de la creciente inquietud que transmite la literatura específica desde la perspectiva profesional (AR@HKU, 2009a; Boud & asociados, 2010; Center for Learning Excellence of University of Bedfordshire- CLE-UB, 2011; CSHE, 2002; Sabirón & Arraiz, 2013), a juicio del alumnado la evitación del fraude y el plagio representa el aspecto menos importante en la praxis evaluativa y también el menos controlado.

Factores de la competencia evaluativa docente que generan dispersión de respuesta en la muestra discente...

La posición en la que sitúa la muestra el resto de factores asociados a la competencia evaluativa docente responde a un mismo patrón. Junto al reconocimiento expreso de su relevancia, el cumplimiento percibido alcanza justo la media teórica y

presenta un elevado índice de dispersión. Este último dato llama a la cautela en la interpretación de las conclusiones asociadas a la *Adecuación y precisión de la evaluación* y la *Planificación evaluativa* y hace prever la convivencia de visiones desiguales sobre las mismas entre el alumnado.

La ***Adecuación y precisión*** representan dos atributos *relevantes* de la evaluación para la muestra discente. Su cumplimiento real suscita entre las/os estudiantes percepciones muy diferentes, provocando en las respuestas un grado de dispersión de más de un punto. Esta circunstancia explicaría que el promedio de valoración resultante presente la *validez* (ítem E.5.1.) y la *fiabilidad* (ítem E.5.2.) como aspectos adecuadamente resueltos en la práctica (satisfechos *bastantes veces*) y que, a la vez, ambos figuren en la lista de áreas de mejora a resolver en la evaluación (objetivo F3).

Las aportaciones procedentes de otros estudios realizados desde la perspectiva discente tampoco son concluyentes al valorar estos atributos. Los resultados relativos a la validez reafirman la imagen global de idoneidad referida, describiendo sistemas evaluativos capaces de reconocer el trabajo y aprendizaje realizado (Buckley, 2012; Rué, 2009b). A la hora de valorar la fiabilidad, en cambio, la investigación próxima pone de manifiesto la misma dispersión que se registra en la muestra analizada. Así por ejemplo, cuestionada la capacidad de la evaluación para discriminar distintos niveles de aprendizaje, Tejedor (1999) y MacLellan (2001) detectaron en sus respectivas muestras importantes niveles de discrepancia entre el alumnado consultado. También se detectan discordancias reseñables en los resultados comunicados en estudios diferentes. Tal es el caso de Jessop (2012) y Buckley (2012) que, ubicados en el mismo contexto geográfico y cronológico, ponen de manifiesto conclusiones distintas en cuanto a la precisión del proceso de calificación. Mientras que el primero describe una relación frágil entre los criterios de evaluación y la adjudicación de calificaciones, el segundo señala esa relación como uno de los aspectos mejor valorados por el alumnado en la evaluación.

De acuerdo con Mateo (2000) o De la Orden (2011b), la complejidad natural de la evaluación del aprendizaje hace que la exploración de la validez y la fiabilidad no se limite a una mera estimación estadística. La precisión o fiabilidad, por ejemplo, está condicionada por la variedad y cantidad de herramientas que se empleen, por la capacidad de medida de las mismas o por las condiciones de administración y la incidencia de sesgos de imprecisión de diversa naturaleza, entre otros factores. La validez, por su parte, guarda relación con la alineación de la evaluación, así como con la selección de demandas adecuadas desde diversas perspectivas (adecuación al perfil discente, a la función evaluativa que se pretenda desempeñar, etc.). Desde esta perspectiva, consideramos que una posible causa de las discrepancias internas y externas evidenciadas en torno a estos atributos en la investigación

puede residir en la perspectiva simplista desde la que se acomete su exploración empírica. En los trabajos referidos en la discusión de este factor, en las escalas de evaluación de la calidad docente analizadas (objetivo D), así como en las propias herramientas empleadas en la presente investigación, el cuestionamiento de la validez y la fiabilidad se viene haciendo por medio de dos preguntas prototípicas, levemente adaptadas: hasta qué punto capta la evaluación aquello que pretende (validez) y con qué grado de precisión discrimina los resultados que genera (fiabilidad). Aunque operativas, estas cuestiones se circunscriben a la visión general de cada atributo, sin captar la diversidad de matices que acompañan a cada uno de ellos. Desde esta premisa, sospechamos que su uso pudiera estar dando lugar a dos consecuencias:

- La emergencia de interpretaciones diferentes para un mismo indicador dentro de la muestra, según la predisposición individual a valorar más un matiz u otro.
- La falta de comprensión o interpretaciones de las diferentes realidades y percepciones que convergen en el promedio.

La **Planificación del proceso evaluativo** hace referencia a cinco indicadores de la competencia evaluativa: la integración de la evaluación en el proceso formativo, su alineación o coherencia, la articulación de los usos pedagógico y acreditativo, la previsión de las herramientas a emplear y la difusión y clarificación de las condiciones anteriores. Valorados todos ellos como *relevantes*, el alumnado consultado da cuenta de situaciones diferentes en cuanto a su manifestación práctica, lo que genera nuevamente altos índices de dispersión.

El principio de la *alineación de la evaluación con los objetivos formativos y la metodología empleada* (ítem C.1.2.) representa uno de los que mayor respaldo recibe en la literatura al definir la calidad evaluativa (Biggs, 2005; Biggs & Tang, 2011; Combs et al., 2008; L. Prieto, 2004). También en la muestra discente, donde figura como el más importante de los indicadores asociados a la competencia evaluativa del profesorado y como una de las fortalezas reconocidas en la evaluación (objetivo F3). A pesar del acuerdo teórico que suscita, según Stiggins (2002) esta coherencia se quebranta a menudo en la práctica, con el planteamiento de situaciones evaluativas desligadas de los objetivos explicitados en la fase de planificación. Esa visión no se corresponde en la muestra estudiada, donde el alumnado advierte un nivel *adecuado* de relación o coherencia, similar al detectado por Tejedor (1999) y Gutiérrez et al. (2013) en sus respectivos trabajos.

La *continuidad de la evaluación* (ítem C.1.1.) representa una cualidad ampliamente atendida en la investigación específica, con desigual resultado. Estudios recientes han

puesto de manifiesto que, frente a los formatos finalistas, los sistemas de evaluación continua dan lugar a un rendimiento académico significativamente superior entre el alumnado (Arribas, 2012; Castejón et al., 2012; Fraile et al., 2013; Holmes, 2015; R. Romero et al., 2014). La continuidad de la evaluación y la realización de pruebas diferentes a lo largo de su desarrollo figura entre los rasgos evaluativos que caracterizan al profesorado mejor valorado por las/os estudiantes (B. Gargallo et al., 2010). La respuesta de la muestra así lo sugiere, ya que presenta la continuidad evaluativa como un indicador relevante a mantener y su mayor fortaleza (objetivo F3).

En cuanto a su manifestación práctica, la continuidad parece estar sujeta a una cierta evolución positiva en los últimos años. Claramente avalada en los programas docentes y en la normativa (Ibarra & Rodríguez, 2010; Padilla et al., 2010), su traslado a la realidad de las aulas está teniendo lugar de manera progresiva, aunque por el momento la evaluación final sigue siendo la más evidente en la práctica (Villa et al., 2013). Ese parece ser el patrón al que responden las valoraciones registradas en la muestra discente, que se sitúan a medio camino entre la evaluación estrictamente final de la que daban cuenta varios estudios antes de emprender el proceso de convergencia (Flórez, 1999; I. González, 2003; Tejedor, 1999; Trillo & Porto, 1999) y las expectativas comunicadas en la teoría evaluativa de referencia (modelo ITC). El alumnado participante reconoce la existencia de tareas evaluativas en las distintas fases del proceso formativo, pero no la adecuada graduación de su intensidad; acusa una *sobrecarga de trabajo en el tramo final del ciclo formativo* y una *descoordinación de las exigencias evaluativas* por parte del profesorado que la agudiza (objetivo F3). Sin hacer alusión expresa a su concentración final, existen otros trabajos que también han acusado la acumulación excesiva de tareas en dispositivos de evaluación continua desde la perspectiva discente (Hamodi & López, 2012; Hortigüela et al., 2015; V. M. López, 2008). Jessop et al. (2014) sugieren además que la sucesión de demandas evaluativas puede llevar a fragmentar artificialmente el aprendizaje y a generar confusión en cuanto a los logros a realizar.

La *combinación de las funciones pedagógicas y acreditativas* (ítem C.1.3.) es el aspecto menos importante de la planificación evaluativa para las/os estudiantes de la muestra. Confirmando las conclusiones de Gutiérrez et al. (2011) y V. M. López y Palacios (2012) para el mismo contexto formativo, su ejecución se sitúa por debajo de la media teórica, poniendo de manifiesto que se intercalan demandas formativas y sumativas, aunque con evidente predominio de las últimas.

El empleo de estrategias de evaluación formativa es el rasgo que mejor caracteriza el estilo evaluativo de las/os buenos docentes según el alumnado (B. Gargallo et al., 2010) y se reconoce como una importante ayuda en el proceso aprendizaje (Fraile et al., 2013;

Hamodi & López, 2012). Aun así, las conclusiones de las que da cuenta la literatura específica confirman el predominio acreditativo evidenciado en la muestra (Banta, 2007; MacLellan, 2001). Incluso en la normativa y en los programas docentes se llega a producir una cierta confusión entre evaluación y calificación (Ibarra & Rodríguez, 2010; Padilla et al., 2010).

Preguntado sobre los *instrumentos* puestos en juego en la evaluación, el alumnado de la muestra admite trabajar con herramientas y situaciones variadas, más orientadas a probar sus competencias que a la mera repetición de contenidos (ítem C.1.4.). De hecho, la combinación de técnicas distintas, más allá del clásico examen, representa la segunda fortaleza más reseñada de la evaluación (objetivo F3), confirmando las observaciones de B. Gargallo et al. (2010) y el CSHE (2002). Sin embargo, estudios realizados en el mismo ámbito formativo revelan resultados que no se corresponden plenamente con esta descripción e introducen matices en dos posibles direcciones:

- En determinados casos se reconoce la diversidad metodológica empleada en la evaluación, pero no su manifestación normalizada, de modo que el alumnado sigue acusando un claro predominio del examen, atendiendo a su frecuencia y a su ponderación en la calificación (Arribas et al. 2010; Gutiérrez et al., 2011; V. M. López & Palacios, 2012; L. F. Martínez et al., 2012; Martínez-Mínguez & Flores, 2014). El análisis de la documentación evaluativa también ofrece reflejo de la instalación progresiva de nuevas técnicas y métodos de evaluación, aunque por el momento sin desbancar la vigencia de los exámenes parciales y finales como primera referencia (Ibarra & Rodríguez, 2010; Padilla et al., 2010).
- En otros casos la discrepancia se produce al detallar el tipo de demanda o tarea, donde se sigue evidenciando el predominio de pruebas basadas en la evocación y el recuerdo, más que en la aplicación. Así lo constatan estudiantes de programas de formación básica de maestras/os y profesorado (Gutiérrez et al., 2013; Ruiz-Gallardo et al., 2013) y de otras titulaciones universitarias (V. Álvarez et al., 2002; Baartman et al., 2013; MacLellan, 2001).

Con todo, la evolución ocurrida en torno a la instrumentación de la evaluación resulta evidente en comparación con las prácticas monometódicas descritas por el alumnado a finales de la década pasada (M. Fernández, 1989; Flórez, 1999). Buena parte de este avance deriva presumiblemente de la implantación del EEES y de la adopción de las competencias como núcleo central de los planes formativos, lo que exige la aplicación de nuevas herramientas de evaluación y de mayor variedad metodológica (Baartman et al., 2006, 2007; Tobón, 2010). A pesar de que las técnicas y herramientas evaluativas hayan

adoptado un protagonismo indiscutible en las primeras acciones formativas para favorecer la convergencia (Bordas & Cabrera, 2001; Knight, 2005; Verdejo, 2008), su instalación definitiva en la práctica exige un proceso más complejo y profundo, y así parece manifestarse en los resultados presentados.

La *clarificación e interiorización* de los criterios definidos para la evaluación (ítem C.2.1.) es el último indicador asociado a la Planificación y muestra una evolución similar a la descrita para los instrumentos o la continuidad. La investigación señala este indicador como uno de los rasgos compartidos por el profesorado universitario mejor valorado por las/os estudiantes (V. Álvarez et al., 1999; Bain, 2007; B. Gargallo et al., 2010) y así se evidencia entre las escalas empleadas para la evaluación de la calidad docente, donde destaca el requisito más reiterado (objetivo D). Estrechamente conectado con el derecho de información que asiste al alumnado en el proceso evaluativo (ENQA, 2009; Real Decreto 1791/2010), su necesidad y pertinencia también se encuentra ampliamente suscrita en la literatura específica del último cuarto de siglo, por contribuir al ajuste de las expectativas de aprendizaje y a procesos de regulación y autorregulación a partir de las mismas (Astin et al., 1992; Boud & asociados, 2010; Brown et al., 1996; J. R. Frederiksen & Collins, 1989; Gibbs & Simpson, 2009; Weston Manor Grupo- WMG, 2007). Sin embargo, consultado el punto de vista discente, su cumplimiento en la práctica se sitúa a medio camino entre dos puntos: no satisface plenamente los requisitos de deseabilidad determinados en el modelo ITC, pero tampoco responde a la imagen de opacidad de la que daban cuenta Flórez (1999) y M. Fernández (1989) en décadas pasadas.

Estudios recientes sobre la satisfacción discente ponen de manifiesto que la claridad de la información evaluativa representa uno de los aspectos mejor valorados en la evaluación (Buckley, 2012; Gil-Flores, 2012; Rué, 2009b). En la muestra analizada también figura como una de las fortalezas reconocidas por el alumnado (objetivo F3), aunque los promedios de respuesta que se registran para su cumplimiento no permiten concluir una satisfacción *adecuada* de las condiciones apuntadas en el modelo ITC. Las/os estudiantes reconocen que las condiciones y criterios que rigen la evaluación se comunican al comienzo del curso y en la documentación básica de la asignatura. Así lo han demostrado también estudios centrados en el análisis de la información evaluativa explicitada en los programas o planificaciones docentes (Padilla et al., 2010; M. A. Sánchez et al., 2010). Sin embargo, cubierto este primer requisito de publicidad, el alumnado afirma que durante el proceso formativo no se llevan a cabo acciones específicas para reforzar la comprensión e interiorización de los criterios definidos.

En contra de las recomendaciones de Price y O'Donovan (2006) o Carless (2002, 2006), la omisión de este segundo nivel de información y comunicación puede dar lugar a

interpretaciones dispares sobre las expectativas de aprendizaje entre el alumnado. Éstas son frecuentes por la falta de herramientas para decodificar los estándares de logro propios de la disciplina (Higgins et al., 2002; Price et al., 2010). Sin embargo, el trabajo específico sobre los criterios de evaluación, una vez concluida su presentación, no tiene lugar en ésta, ni en otras muestras analizadas (Jessop, 2012; G. Rodríguez et al., 2012).

En contra de la percepción de Ibarra y Rodríguez (2010), que señalan el cumplimiento generalizado de la transparencia evaluativa, al menos al nivel informativo, la investigación próxima da cuenta de casos en los que ni siquiera ese primer requisito se considera cubierto. Tanto en la formación básica de maestras/os (Arribas et al., 2010; Gutiérrez et al., 2011; V. M. López & Palacios, 2012), como en otras titulaciones universitarias (Burgess, 2004; Tejedor, 1999), el alumnado afirma no disponer de suficiente información en cuestiones elementales vinculadas a los dispositivos evaluativos (criterios de calificación, fechas y plazos, etc.).

La atención a la diversidad en la evaluación desde la perspectiva de la muestra discente...

La revisión de la percepción discente sobre los factores asociados a la competencia evaluativa docente no podía concluir sin hacer referencia a la atención a la diversidad. Según ha sido justificado en el capítulo séptimo, los dos ítems referidos a la adecuación de los dispositivos evaluativos (ítems C.3.1. y C.3.2.) han sido excluidos del cuestionario discente por razones psicométricas. Aunque los argumentos que avalan tal decisión son de naturaleza estadística, su valoración debe llevarse a cabo desde otros ángulos y a ello responde esta reflexión final.

Para que un indicador se ubique claramente en la estructura factorial, es preciso que la muestra tienda a situarlo en el mismo lugar de forma reiterada. Sin embargo, al igual que en la muestra docente, entre el alumnado se constatan claros indicios del desacuerdo que suscita la atención a la diversidad al aproximarse al terreno de la evaluación. Como en la normativa evaluativa, las/os estudiantes consultadas/os parecen entender la justicia en términos de igualdad y así se refleja en las fortalezas y debilidades que apuntan en la praxis evaluativa (objetivo F): mientras que las diferencias se reseñan como *agravios comparativos* en la lista de aspectos a modificar, la *igualdad en las condiciones* evaluativas figura entre los aspectos positivos a mantener. En consecuencia, barajamos que los riesgos psicométricos que han aconsejado la supresión de esta dimensión puedan responder al desacuerdo real que se produce en torno a la misma, más que a su ambigüedad.

2.2 Relación entre variables socio-demográficas, académicas y profesionales y la competencia evaluativa docente percibida por el alumnado: conclusiones y reflejo de las mismas en la investigación y la literatura (Objetivo F2).

La visión general descrita en la muestra discente difiere en función de una amplia gama de variables socio-demográficas y académico-institucionales. Aunque abundantes y significativas, las relaciones y diferencias detectadas presentan en su mayoría un efecto de magnitud pequeña, especialmente en las valoraciones de importancia, donde se produce mayor consenso entre el alumnado. Sólo en dos de las variables analizadas se constatan diferencias con un tamaño de efecto grande. Dichas variables son la competencia docente percibida y la materia, que dan lugar a percepciones significativamente diferentes del cumplimiento de la competencia evaluativa entre el alumnado.

Según los datos resumidos en el capítulo anterior, las/os estudiantes perciben prácticas evaluativas significativamente diferentes en docentes considerados competentes, poco competentes y muy competentes. Éstas afectan a todos los factores ligados a la competencia evaluativa y responden a un mismo patrón: las puntuaciones mayores de realización se registran entre el alumnado que mayor **competencia docente** reconoce al profesorado. Esta tendencia se confirma en el hallazgo de una correlación, muy significativa y de magnitud pequeña-moderada, entre la competencia docente percibida por las/os estudiantes y el grado de realización estimado en los factores involucrados en la competencia evaluativa. Las propias limitaciones del diseño de la investigación, no permiten determinar la direccionalidad de esta relación. En cualquier caso, la unión entre la competencia docente y la práctica evaluativa pone en evidencia que la evaluación del aprendizaje representa para el alumnado una cuestión relevante, tal y como se señala desde la literatura (Biggs, 2005; Dochy et al., 2002; Gibbs & Dunbar-Goddet, 2007; Morales, 2008c). Tanto por su efecto potencial en el desarrollo profesional docente, como por su posible influencia en los juicios de valor que emite el alumnado al respecto, esta asociación proporciona un argumento en favor de quienes advierten en la evaluación del aprendizaje un factor estratégico de la calidad institucional (Gibbs, 2003, 2010; Kaneko, 2008; Porto, 2010; Stiggins & Duke, 2008).

En las investigaciones próximas no hemos hallado datos directos que confirmen o refuten la relación evidenciada. La aportación más próxima la proporciona Buckley (2012), que evidenció en el contexto del Reino Unido correlaciones significativas y positivas entre la calidad de la evaluación y el feedback recibido y dos variables relevantes: el apoyo académico percibido y la satisfacción general manifestada con la formación.

En la formación básica de maestras/os de Educación Infantil y Primaria el módulo o asignatura en la que se haya llevado a cabo la valoración se inscribe en una determinada **materia**, según lo dispuesto en las órdenes de referencia (ECI/3854/2007 y ECI/3857/2007, respectivamente). Considerando las materias comunes y específicas que se regulan para ambas titulaciones, esta variable discrimina en la muestra 24 valores o grupos diferentes. Las valoraciones que se registran en los mismos difieren significativamente en todos los factores analizados, con mayor efecto en las estimaciones de realización que en las de importancia.

La reducción natural de las materias en ambas titulaciones, tiene lugar por medio de cuatro **módulos** en torno a los que se ordenan las diversas asignaturas (formación básica, didáctico disciplinar, mención y DECA). Replicada la comparación, persisten las diferencias manifestadas en el cumplimiento atribuido por materia, pero con efecto de magnitud pequeña. Aunque no respondan a un patrón regular, en ellas se advierte un especial protagonismo del alumnado del módulo de formación básica, único grupo implicado en todas las diferencias significativas identificadas.

La referencia a las materias y módulos no se contempla expresamente en investigaciones previas efectuadas sobre el tema. El área o rama de conocimiento representa la referencia más próxima, pero su análisis no da lugar a resultados concluyentes, como ya ha sido advertido al presentar las conclusiones relativas a la visión docente (objetivo E2). Por tanto, no contamos con una explicación certera sobre las diferencias detectadas en función de la materia y el módulo de referencia, aunque a la luz de las conclusiones obtenidas para el objetivo E2, advertimos ciertos indicios que nos llevan a sugerir una posible relación de las mismas con la formación previa del profesorado:

- La realización percibida en los factores asociados a la competencia evaluativa en función del módulo de referencia ha dado lugar a siete diferencias significativas dentro de la muestra; en cinco de ellas, el promedio superior de respuesta se registra entre el alumnado que ha efectuado su valoración para asignaturas del módulo de formación básica.
- De lo anterior se desprende que las/os estudiantes tienden a percibir un mejor desempeño evaluativo en el módulo de formación básica, aquel en el que se produce una mayor concentración de profesorado titulado en Psicología, Pedagogía o Psicopedagogía. Para indagar sobre esta presunción, se ha llevado a cabo un nuevo análisis subdividiendo la muestra discente en dos grupos: estudiantes que han efectuado sus valoraciones para sistemas evaluativos de profesorado de Psicología, Pedagogía o Psicopedagogía y estudiantes que han valorado la

competencia evaluativa de docentes de otras titulaciones. Como resume el **ANEXO 30**, el primer grupo percibe un cumplimiento significativamente superior en todos los factores considerados, salvo en uno (Regulación del aprendizaje por medio de la evaluación).

- Las diferencias advertidas en las valoraciones de la muestra discente en función del curso también pueden ser interpretadas desde esta apreciación. El profesorado titulado en Pedagogía, Psicología o Psicopedagogía se concentra principalmente en el primer curso de la titulación, aquel en el que las/os estudiantes conceden mayores promedios de realización al valorar los factores de la competencia evaluativa.

V. Álvarez et al. (2002) y Gil-Flores (2012) señalan que dentro de las instituciones formativas se generan realidades diferenciales y subculturas específicas en función del área de conocimiento. Los indicios presentados así lo sugieren, dejando entrever modos diferentes de llevar a cabo la evaluación en función de la materia y/o la formación previa del profesorado.

2.3 Obstáculos y apoyos en el ejercicio de la evaluación según la muestra discente y su reflejo en la investigación y la literatura (Objetivo F3).

Para completar el diagnóstico, el objetivo F3 se ha detenido en la detección de las principales fortalezas y debilidades percibidas por el alumnado en la evaluación de su aprendizaje. Las respuestas de la muestra a dos preguntas abiertas devuelven una visión más bien negativa de la evaluación, donde las áreas de mejora sugeridas exceden la recomendación de aspectos a mantener. El grado de coincidencia detectado entre las aportaciones discentes y los indicadores abordados en el cuestionario sugiere una cierta transformación de la muestra (Manzano & Braña, 2005).

Las fortalezas y debilidades que las/os estudiantes atribuyen a la evaluación del aprendizaje guardan una estrecha conexión con dos de los factores asociados a la competencia evaluativa docente: la Planificación del proceso evaluativo y la Regulación del aprendizaje.

- Las deficiencias asociadas al feedback representan la mayor dificultad evaluativa en opinión del alumnado, reafirmando las conclusiones de otros muchos estudios

efectuados sobre el tema (Arribas et. al., 2010; Ferguson, 2011; Gutiérrez et al., 2011, 2013; Jessop, 2012; L. F. Martínez et al., 2012; G. Rodríguez et al., 2012).

- Probablemente por influencia de los cambios pedagógicos y estructurales sujetos al EEES, para la muestra discente las mayores fortalezas evaluativas son la continuidad, la variedad metodológica aplicada en su desarrollo y la claridad o publicidad en sus condiciones de partida. Así se ha recogido también en otros trabajos recientes (Buscà, Pintor et al., 2010; Ibarra & Rodríguez, 2010; Padilla et al., 2010; M. A. Sánchez et al., 2010).

Partiendo de esa base, a continuación se detallan aquellas otras aportaciones con menor incidencia en las respuestas recopiladas y su reflejo en otras investigaciones.

- Reafirmando las inquietudes que comunican el AR@HKU (2009a) y el CSHE (2002, 2010), el trabajo en grupo se manifiesta como un objeto evaluativo polémico. Aparece referido simultáneamente en la lista de fortalezas y en la de aspectos a mejorar. A favor se presenta el argumento del equipo como espacio para el apoyo mutuo y el aprendizaje, pero a la vez se señalan tensiones asociadas al reparto de la carga de trabajo y su reflejo en la calificación. La holgazanería social (Latané, Willians, & Harking, 1979) representa la manifestación más clara de dicha dificultad. Esta valoración intermedia por parte de las/os estudiantes también se respalda desde el punto de vista del profesorado, que acusa dificultades para valorar el producto de trabajos efectuados en grupo en la muestra analizada (objetivo E3), así como en otras aproximaciones empíricas o teóricas efectuadas al tema (Chan, 2010; Heathfield, 2003; Iranzo, Gimeno, & Sánchez, 2012; Morales, 2008b; Rizo, 2004).
- Relacionada con el volumen de trabajo demandado y con la dimensión que va de la exigencia a la indulgencia, la accesibilidad de la asignatura representa una de las cuestiones más discutidas al valorar la fiabilidad del alumnado como agente evaluador (Brodie, 1998; Marsh, 1987; Richardson, 2005). Desde esta perspectiva, B. Gargallo et al. (2010) concluyeron en su estudio que más de la mitad del alumnado consultado presentaba la accesibilidad como rasgo evaluativo deseable. Sin embargo, en la muestra analizada la facilidad sólo figura entre las fortalezas citadas y de forma minoritaria.
- Como en algunos de los antecedentes consultados (B. Gargallo et al., 2010; Tejedor, 1999), el alumnado ve con buenos ojos la inclusión de la asistencia y participación en clase como componentes ponderados en la calificación, aunque se trata de una fortaleza escasamente representa en sus observaciones. Ante ello, suscribimos la

crítica que realizan Sabirón y Arraiz (2013) a la valoración acreditativa de la asistencia, por considerar que transmite un mensaje erróneo sobre la simplicidad del aprendizaje, más basado en la recepción que en la construcción.

3. Objetivo G. Contrastar el punto de vista que manifiestan docentes y discentes con respecto a la competencia evaluativa del profesorado vinculado a la formación básica de maestras/os en la UD y BAM, indagando tanto en la perspectiva teórica (importancia), como en la práctica (realización)

Los objetivos E y F han analizado en profundidad la visión que manifiestan profesorado y alumnado sobre la competencia evaluativa docente desde su particular perspectiva. Admitido que la diferencia en roles, funciones o bagaje formativo provoca percepciones desiguales sobre una misma realidad en docentes y discentes (V. Álvarez et al., 2002), el tercer objetivo del marco empírico plantea el contraste explícito de las respuestas de ambos sectores de la muestra en torno a los nueve indicadores comunes asociados a la competencia evaluativa y las cuatro dimensiones en las que se organizan.

La primera conclusión que arroja el contraste de visiones es que ambas fuentes reconocen un claro desajuste entre la importancia y el cumplimiento atribuido a los indicadores y dimensiones comunes de la competencia evaluativa, situando todos ellos en el área de Concentración de la rejilla AIR. Sin embargo, su respuesta difiere de acuerdo a una tendencia que ya ha sido ampliamente informada en estudios similares (Gutiérrez et al., 2011, 2013; Martínez-Mínguez & Flores, 2014; Mas, 2012; Ruiz-Gallardo et al., 2013; Tejedor, 1999). Ésta se caracteriza por la manifestación de una postura más crítica y exigente entre el alumnado quien, en comparación con la muestra docente, ofrece mayores estimaciones al valorar la importancia y puntuaciones menores al juzgar el grado de realización.

Para profundizar en este patrón recurrente de respuesta, Feldman (1988) llevó a cabo un trabajo de revisión de varios estudios basados en el contraste de las perspectivas docente y discente. Su principal conclusión fue que en las respuestas se produce uniformidad intra-grupal y divergencia inter-grupo. Así se constata en la presente

investigación, aunque con ciertos matices respecto a la uniformidad dentro de la muestra docente y discente⁹⁹. En cuanto al origen de dicha divergencia, Gutiérrez et al. (2013) y Ruiz-Gallardo et al. (2013) señalan la confluencia de dos tipos de sesgo como causa probable: la deseabilidad social y la negatividad. El primero afectaría prioritariamente al profesorado, provocando la sobrevaloración del cumplimiento de los indicadores cuestionados, y el segundo al alumnado, predispuesto a su infravaloración por la acumulación de vivencias y experiencias negativas con relación a la evaluación (L. F. Martínez et al., 2012; Ricoy & Fernández-Rodríguez, 2013; Ureña et al., 2009).

B. García et al. (2012) sugieren que, salvada esta tendencia general de respuesta, docentes y estudiantes coinciden en las principales fortalezas y debilidades atribuidas en sus valoraciones. Así se produce en la presente investigación, donde alumnado y profesorado revelan en sus valoraciones una lógica común de priorización, con acuerdos evidentes sobre los aspectos que mejor y peor se desarrollan en la praxis evaluativa.

La escasa incidencia de la **Participación discente en la evaluación** (ítem C.7.1.) representa uno de los acuerdos más evidentes en ambos sectores de la muestra. Tanto para estudiantes como para profesoras/es, las fórmulas participativas de evaluación constituyen el indicador menos relevante y el menos llevado a cabo, configurando un área de mejora evidente, pero no prioritaria. Ambas fuentes coinciden en describir una intervención discente puntual, centrada en la corrección de trabajos propios y ajenos durante el proceso formativo. Aunque muy pequeña en magnitud y efecto, la diferencia de puntuaciones registrada para describir su realización resulta estadísticamente significativa.

Esta tónica general de acuerdo en torno a la escasa incidencia de las fórmulas de evaluación participativa también se encuentra reportada en otras investigaciones similares. Contrastando el punto de vista de estudiantes y responsables de departamentos (Porto, 2006), estudiantes y profesorado (Gutiérrez et al., 2011) o estudiantes, alumnado egresado y profesorado (Arribas et al., 2010; Gutiérrez et al., 2013) la conclusión compartida apunta a la preponderancia de formatos de heteroevaluación interna a cargo docente. Las pocas discrepancias que se constatan con respecto al tema hacen referencia a la importancia relativa de la intervención discente en la evaluación o a las actitudes que genera su puesta en práctica. G. Rodríguez et al. (2012), por ejemplo, detectaron diferencias entre

⁹⁹ Tanto alumnado como profesorado informan de sus apreciaciones con respecto a dispositivos evaluativos distintos, de modo que sus valoraciones difieren al no responder a un objeto común, particularmente en la subescala de realización. Aunque los resultados no hayan sido empleados a los efectos de la presente investigación, sí se ha constatado una dispersión mucho menor al comparar las valoraciones de las/os estudiantes para un mismo dispositivo evaluativo. La divergencia entre las valoraciones docente y discente, en cambio, se siguen apreciando incluso cuando refieren al mismo objeto exacto.

estudiantes y profesorado al valorar la pertinencia y las oportunidades de intervención discente en el diseño de la evaluación. C. Brew et al. (2009) por su parte, revelaron motivaciones desiguales para implicarse en experiencias de coevaluación o autoevaluación, pedagógicas y formativas en el caso del alumnado y fundamentalmente prácticas desde la perspectiva docente.

Frente a la baja prioridad que merece para la muestra la evaluación participativa, alumnado y profesorado coinciden en situar la **Regulación del aprendizaje** como la más importante de las dimensiones vinculadas a la competencia evaluativa. Según la tendencia de respuesta anticipada, el alumnado muestra valoraciones significativamente superiores al calificar la importancia de la *comunicación de los resultados de aprendizaje* (ítem C.5.3.) y de la propuesta de *pautas de proacción para su mejora* (ítem C.6.2.). Los mayores desacuerdos se producen en sentido inverso y al juzgar el reflejo que tiene en la práctica dicha regulación. Tanto en la *comunicación de resultados* (ítem C.5.3.), como en la *comisión de feedback* (ítem C.6.1.) y de *pautas de actuación* (ítem C.6.2.), estudiantes y profesorado manifiestan índices de cumplimiento significativamente diferentes, igual que en otros estudios realizados en el contexto próximo. Las principales discordancias reportadas tienen que ver con la frecuencia de la comunicación de los resultados y los comentarios derivados, siempre más presentes según docentes (Arribas et al., 2010; Gutiérrez et al., 2013; L. F. Martínez et al., 2012; Martínez-Mínguez & Flores, 2014). Según la cuantificación de Gutiérrez et al. (2011) estos comentarios son frecuentes para el 78% del profesorado y sólo para el 17% del alumnado. Los hallazgos de G. Rodríguez et al. (2012) revelan una situación similar en cuanto a las orientaciones prospectivas para mejorar el trabajo y el aprendizaje, ofrecidas según el 82% de la muestra docente y reconocidas por el 32% de las/os estudiantes.

Admitidas como rasgo general, una de las explicaciones que arroja la literatura para explicar tales discrepancias son las diferentes concepciones que sostienen del feedback docentes y discentes (Boud & Molloy, 2013; Cano, 2014; Carless, 2006; Higgins et al., 2002; Orsmond & Merry, 2011; Price et al., 2010; Sambell, 2011). El alumnado identifica el feedback con informes y comentarios escritos que proporcionan la respuesta correcta, mientras que el profesorado interpreta que en el desarrollo de las actividades cotidianas también se produce retroalimentación (Sambell et al., 2007). Este cruce de concepciones provocaría que toda retroalimentación que no se ajuste a dicho patrón pase inadvertida, incluso habiendo sido deliberadamente emitida por el profesorado para orientar el cambio.

Interpretados en su conjunto, estos datos sobre la alta importancia y la insuficiente realización de los procesos de retroalimentación y proacción dejan abierta una línea de mejora prioritaria en el ámbito evaluativo. Tanto si se adopta el punto de vista más crítico

(estudiantes) como el más indulgente (docentes), la descripción de este factor en la práctica pone de manifiesto una distancia considerable con respecto a las pautas de deseabilidad definidas en el modelo ITC. A diferencia de lo que se constata para la evaluación participativa, la Regulación del aprendizaje por medio de la evaluación representa para ambos sectores una inquietud prioritaria y compartida, y así se ha constatado en las áreas de mejora recogidas en cada caso (objetivos E3 y F3). Todo ello nos lleva a suscribir la conclusión que alcanza Evans (2013) tras una extensa revisión de artículos publicados sobre el tema entre 2000 y 2012: aunque los principios de calidad asociados al feedback resulten claros y compartidos, su puesta en práctica sigue siendo problemática. Como ha sido advertido al analizar la perspectiva docente, para el profesorado estas dificultades guardan conexión directa con la sobrecarga que generan las condiciones laborales y el número de estudiantes (objetivo E3).

El reconocimiento de la **Planificación estratégica de la evaluación** como la dimensión más llevada a cabo de la competencia evaluativa docente también es objeto de acuerdo entre ambos sectores de la muestra. Ésta incluye cuatro indicadores asociados al diseño de los dispositivos evaluativos y la visión de estudiantes y profesorado sólo difiere significativamente al valorar la *continuidad de la evaluación* (ítem C.1.1.). Desde el punto de vista del profesorado, la evaluación se desarrolla por medio de tareas en las tres fases del proceso, aunque la intensidad de las mismas no llega a graduarse ni coordinarse adecuadamente. El alumnado manifiesta un punto de vista más crítico al respecto, sin reconocer apenas la incidencia de la evaluación inicial, procesual y final en su proceso formativo. Los desacuerdos a la hora de juzgar el grado de continuidad de la evaluación en el proceso de enseñanza y aprendizaje también aparecen en otros estudios centrados en el análisis de las praxis evaluativa, tanto en la formación básica de profesorado (Arribas et al., 2010) como en otras titulaciones (Porto, 2006).

Los datos evidenciados en el presente trabajo y los recopilados de la investigación próxima sugieren que ambos sectores de la muestra pudieran estar realizando sus valoraciones desde concepciones diferentes de la continuidad evaluativa. Retomando la discusión iniciada al analizar el punto de vista docente (objetivo E1), tras la divergencia de opiniones que se produce, intuimos dos aproximaciones alternativas para caracterizar el grado de continuidad de los dispositivos evaluativos:

- La primera de ellas es la que podríamos caracterizar como *instrumental*, centrada en el número de pruebas planteadas para la valoración del aprendizaje y en la ubicación de las mismas en las distintas fases del proceso formativo. Esta interpretación podría dar lugar a la visión más optimista constatada entre el profesorado a la hora de juzgar la continuidad de la evaluación. Así lo sugieren Villa

et al. (2013) en un estudio reciente, al comparar la perspectiva de docentes y de responsables académicos.

- La segunda parece adoptar una orientación más *pedagógica*, centrada en el impacto formativo de la sucesión de tareas evaluativas en la construcción del aprendizaje. En línea con las aportaciones que se desprenden de la literatura en favor de una evaluación sensible (Brown et al., 1996; Gibbs & Simpson, 2009; Knight, 2005; Monereo, 2003), las/os estudiantes de la muestra sitúan la continuidad como una fortaleza de la evaluación y, a la vez, comunican su inquietud con respecto a la intensidad y al valor didáctico de la misma (objetivo F3).

El contraste de perspectivas finaliza con el indicador que hace referencia a la **claridad de los criterios y condiciones de la evaluación** (ítem C. 2.1.), el segundo más llevado a cabo en opinión de ambas fuentes. Siguiendo la tendencia general de respuesta evidenciada en el contraste, el profesorado reconoce para el mismo un desarrollo práctico significativamente superior del que declara el alumnado. Otros trabajos de investigación también han dejado constancia de esta discrepancia a la hora de valorar la difusión de los criterios de evaluación y calificación. Gutiérrez et al. (2011) por ejemplo, hallaron posiciones muy enfrentadas dentro de su muestra, donde el 85% del profesorado declaraba explicitar los criterios de evaluación al comienzo del curso y el 74% del alumnado afirmaba no tener conocimiento de ello. De modo similar, y en el contexto específico de la formación básica de maestras/os, Arribas et al. (2010) detectaron que la proporción de alumnado actual y egresado que consideraba cubierta esta explicitación rondaba el 20%, muy por debajo del 89% evidenciado en el sector docente. La comunicación sobre el tipo de pruebas a efectuar o la advertencia anticipada de fechas y plazos también representa un foco de discrepancia según los resultados reportados en L.F. Martínez et al. (2012) o Tejedor (1999).

Como en Martínez-Mínguez & Flores (2014), la diferencia de visiones detectada en la muestra tiene lugar en la misma dirección, pero no de forma tan extrema. De acuerdo con la evolución que anuncian para la transparencia evaluativa Padilla et al. (2010) e Ibarra y Rodríguez (2010), alumnado y profesorado coinciden en señalar la satisfacción de un primer nivel informativo, caracterizado por la explicitación escrita de los criterios y condiciones de la evaluación al comienzo del curso. Las discrepancias se producen al cuestionar el desarrollo de acciones específicas para profundizar en su comprensión en el transcurso del proceso formativo. En línea con las recomendaciones que se desprenden de la literatura (Carless, 2002, 2006; J. Chappuis, 2005; Price & O'Donovan; Sadler, 2010, 2013), el profesorado afirma llevar a cabo actividades específicas para recordar, aclarar y

profundizar en el significado de los criterios establecidos, aunque sin recurrir al uso de herramientas específicas como rúbricas o listas de control. La respuesta discente, en cambio, sólo reconoce la explicitación inicial contemplada en la normativa (ENQA, 2009; Real Decreto 1791/2010).

El resto de indicadores cuestionados en ambos sectores de la muestra no resultan directamente comparables, al haber sido planteados con matices diferentes en los respectivos cuestionarios (**ANEXO 17**). No obstante, el contraste indirecto de las visiones reveladas en los objetivos E1 y F1, sugiere una cierta coincidencia en la apreciación general:

- Ambas fuentes coinciden en señalar la escasa incidencia de la metaevaluación en la praxis evaluativa, mostrando el profesorado una postura ligeramente más crítica al respecto.
- Al valorar la coherencia en la puesta en práctica de los sistemas de evaluación diseñados, docentes y discentes sostienen una postura similar, indicando que los ajustes y adecuaciones representan la excepción a la norma.
- La adecuación y precisión de la evaluación figura entre las áreas de mejora a solventar en ambos sectores de la muestra, en especial, en lo que refiere al segundo atributo. El modelo ITC establece que la calificación precisa es aquella que discrimina distintos niveles de consecución en los aprendizajes previstos, según el logro evidenciado en los criterios de evaluación y calificación definidos en la fase de diseño. Se trata de uno de los aspectos en los que menos profundizan los estudios consultados y las conclusiones publicadas muestran diferencias destacables dependiendo de la fuente de la que procedan. A diferencia de lo evidenciado en la muestra, por lo general, el profesorado considera que los procedimientos evaluativos aplicados discriminan correctamente si se han alcanzado o no los objetivos previstos (Serrano, 2010; Tejedor, 1999). En cambio, al consultar el punto de vista discente, coexisten los testimonios de quienes estiman un grado de precisión razonable en la valoración del aprendizaje (Buckley, 2012; Rué, 2009b), y quienes lamentan una falta de rigor y credibilidad (Bryson et al., 2008; Jessop, 2012; McLellan, 2001). El alumnado consultado se atiene a esta última postura, aunque con una visión algo menos crítica que la comunicada por el profesorado.

4. Objetivo H. Describir las experiencias y demandas que plantea el profesorado con respecto a la formación permanente vinculada a la evaluación del aprendizaje

El último de los propósitos vinculado al marco empírico de la investigación centra su atención en la formación permanente de la muestra docente en materia evaluativa. Su análisis incorpora la descripción de las actividades formativas llevadas a cabo por el profesorado (objetivo H1) y de aquellas otras que considera convenientes o necesarias (objetivo H2).

La discusión entablada en torno al objetivo E ha puesto de manifiesto que la formación del profesorado representa un factor clave en el desarrollo de su competencia evaluativa. Como dice Fullan (1999) la formación es al tiempo el problema y la solución, y así se ha constatado en la muestra estudiada, donde figura simultáneamente entre las dificultades y los apoyos detectados para el ejercicio de la evaluación.

La formación previa realizada por maestras/os para el desempeño de las funciones evaluativas en la escuela a menudo se ha calificado de insuficiente (Mertler, 2009; Sanders & Vogel, 1993/2012; Stiggins, 1993/2012, 1998; Wise & Lukin, 1993/2012) e inadecuada (Gullickson, 1993/2012; Stiggins & Conklin, 1988). Esta carencia se hace más evidente en los niveles universitarios, donde el acceso a la enseñanza no está supeditado a la realización de una formación didáctico-pedagógica específica que capacite para ejercer la docencia y, en particular, la evaluación (Banta, 2007; Erwin, 2003; Monereo & Castelló, 2009a).

La investigación y la literatura se hacen eco del riesgo y las repercusiones de esta escasa capacitación evaluativa desde perspectivas:

- En ocasiones, se apunta a ella como principal causa de las disfuncionalidades evaluativas detectadas en la universidad (Flórez, 1999; Nieto, 2000; Porto, 2005a, 2005b; Trillo & Porto, 1999).
- Otros trabajos advierten en la falta de formación un factor decisivo para contribuir a la reproducción de los clásicos patrones evaluativos y la instalación de rutinas irreflexivas al respecto (Clipa & Duta, 2010; Palacios & López, 2013; Tejedor, 1999).
- Cabría valorar asimismo, hasta qué punto incide la falta de preparación específica en las lagunas y la confusión conceptual reportada entre el profesorado universitario

en materia evaluativa (B. Álvarez et al., 2007; Ibarra & Rodríguez, 2010; A. Pérez et al., 2009; Rué, 2008).

Según Flemming (2003) los programas de capacitación inicial en materia evaluativa estarían sujetos a una mayor probabilidad de éxito por el menor arraigo de determinados hábitos y rutinas entre las/os destinatarias/os. Sin embargo, en las condiciones de acceso a la docencia descritas, la formación permanente para la evaluación del aprendizaje adopta especial protagonismo (Caballeros, 2013; Palacios & López, 2013). Ante las deficiencias evaluativas reportadas, diversos estudios coinciden en señalarla como principal sugerencia de mejora (V. Álvarez et al., 2002; Buendía & López, 1999; Burgess, 2007; B. Gargallo et al., 2010; Ibarra & Rodríguez, 2010; Tejedor, 1999). Para Ion y Cano (2011), Mateo (2012) o Torra et al. (2013), debería constituir un foco de atención y un área de mejora institucional, no una necesidad o inquietud propia del profesorado. Sin embargo, estas declaraciones generalizadas en favor de la formación permanente en materia evaluativa, no se corresponden con su realización efectiva como se recoge a continuación.

El testimonio docente sobre la formación permanente llevada a cabo en materia evaluativa...

Respaldando la relevancia y pertinencia reseñada en los trabajos anteriores, algo más del 81% profesorado de la muestra declara sentir ***necesidades formativas relacionadas con su competencia evaluativa*** (objetivo H2). Según Hamodi y López (2012) esta necesidad se encuentra suscrita incluso entre el alumnado egresado, quien incorporado al mercado laboral, reconoce la conveniencia de seguir profundizando en el tema. La implantación del EEES y los cambios pedagógicos y estructurales sujetos a ella han dado lugar a su intensificación (M. González & Raposo, 2009; Ibarra & Rodríguez, 2010; Villa et al., 2013). Sin embargo, la consulta realizada en la presente investigación revela que apenas la cuarta parte del profesorado (23,5%) ha llevado a cabo alguna actividad formativa sobre evaluación, porcentaje similar al que evidenció Tejedor (1999) hace más de una década (18,5%).

Sin desestimar la influencia de los recursos y medios disponibles en este resultado, otros datos sugieren que el interés real del profesorado por la formación evaluativa específica no es tan importante como hacen pensar sus manifestaciones iniciales. De acuerdo con Gutiérrez et al. (2013) el profesorado atribuye a factores estructurales sus dificultades evaluativas, no a la falta de formación. En consecuencia, la evaluación no se

encuentra entre los temas en los que mayor necesidad de dominio manifiesta el profesorado, ni tampoco entre sus prioridades formativas (V. Álvarez et al., 2011; Elexpuru et al., 2006; JCSEE, 2007; Mas, 2012).

Independientemente de las causas que expliquen la diferencia, esa pequeña parte del profesorado que ha realizado formación evaluativa en los dos últimos cursos, lo ha hecho de **forma voluntaria**, confirmando las apreciaciones generales que se desprenden de la literatura (Michavila, 2005; Santos, 2010a; Zabalza, 2002). Los datos recopilados revelan un nivel de **satisfacción razonable** al concluir la experiencia, calificando entre *moderadamente útiles* y *muy útiles* las aportaciones resultantes. Sin embargo, no disponemos de datos o testimonios precisos que permitan realizar una valoración de la transferencia al aula de dicha formación, análisis pertinente y relevante según Prebble et al., 2004; Monereo (2013) o Feixas et al. (2013).

Como anuncia Montero (2006), en el repertorio de **temas específicos tratados** se hace evidente la influencia del planteamiento formativo del EEES y sus repercusiones en la evaluación. Destacan así, las actividades formativas orientadas a la *evaluación de competencias* o la aplicación de las *tecnologías en el proceso evaluativo*. En un plano más teórico, se reseñan también experiencias formativas relacionadas con los *principios generales de la evaluación*, los *instrumentos y técnicas* aplicables en la misma y los posibles *enfoques* para ejercerla.

En cuanto a los **formatos** más extendidos, confirmando las apreciaciones de Caballero (2013) sobre la escasa contribución de los departamentos a la formación docente, el profesorado de la muestra afirma haber indagado en dichos temas por medio de la autoformación, esto es, buscando proactivamente respuesta a sus necesidades e inquietudes en lecturas, jornadas o en la experiencia de otras/os colegas. A pesar de que en el objetivo E2 haya sido descartada su influencia en la competencia evaluativa, la formación de postgrado representa la segunda vía más frecuente para profundizar en la evaluación según declaran las/os participantes. El clásico formato curso, presencial, predefinido y externamente dirigido, aparece con menor frecuencia, pero sigue representando una alternativa destacada para la formación permanente. Pozo et al. (2006) o Marín y Cervi (2006) advierten en este último formato una efectividad potencial limitada para el cambio de las concepciones evaluativas. Su punto de partida es la linealidad de la relación entre teoría y práctica, donde la exposición de la primera se considera suficiente para modificar la segunda (V. M. López, 2004; A. I. Pérez, 2010b; Pozo, 2001). Sin embargo, para incidir en el modo de evaluar es preciso alterar las concepciones de base, lo que exige todo un proceso de reconstrucción y replanteamiento que modifique realmente las ideas y prácticas previas, para emprender la construcción de nuevas alternativas (Marín &

Cervi, 2006; Rodrigo, 1997). Este reto resulta complejo cuando el profesorado adopta el rol de receptor pasivo de formación. La dificultad incrementa en actividades de **corta duración** como las que refiere la muestra, ya que en ellas la probabilidad de poder realizar un contraste práctico disminuye (Prebble et al., 2004; Southwell & Morgan, 2009). Como afirman I. Fernández y Márquez (2014) al respecto, una actividad formativa puntual puede ser una buena estrategia para el desarrollo de aprendizajes instrumentales, pero difícilmente puede afectar a las concepciones profundas.

El testimonio docente sobre la formación permanente demandada en materia evaluativa...

Las **modalidades formativas** sugeridas por el profesorado al describir la formación evaluativa demandada (objetivo H2) se sitúan más próximas a los requisitos detallados en el capítulo tercero para el cambio de concepciones evaluativas profundas. Confirmando las apreciaciones de Margalef y Álvarez (2005) o Caballero (2013), todas ellas parten de la intervención docente directa y la discusión y reflexión colegiada. No se produce ninguna sugerencia en favor de los clásicos cursos y, en su lugar, el *intercambio de experiencias con colegas* destaca como la alternativa formativa predilecta. Consiste en la creación de espacios para el diálogo y el intercambio sobre problemas e inquietudes comunes y sus beneficios se encuentran ampliamente reconocidos (Blanch & Corcelles, 2014; Mateo, 2012; Serafini, 2000; Torra et al., 2013; Villa & García, 2014). Según Bolarín et al. (2013) los procesos de coordinación docente vertical u horizontal pueden constituir espacios dinamizadores de este intercambio. Tanto en su versión presencial directa, como por medio de redes en línea, esta modalidad formativa aparece recogida como sugerencia de mejora en otras investigaciones recientes sobre evaluación o docencia universitaria (Capllonch & Buscà, 2012; B. Gargallo et al., 2010; Ibarra & Rodríguez, 2010; Olmos & Rodríguez, 2010; Tejada, 2009). De hecho, para el profesorado consultado resulta el mayor apoyo en el desarrollo de sus funciones evaluativas (objetivo E3). En ocasiones se le asocia un cierto riesgo de dispersión y desestructuración por exceso de espontaneidad (Prebble et al., 2004), pero puede ser contrarrestado complementándola con otros cauces o sistematizando su desarrollo.

Como en el estudio de Tejedor (1999), el profesorado participante muestra predilección por las acciones formativas basadas en el intercambio entre colegas, pero no desestima la intervención de agentes expertos externos. Más que a la impartición de cursos teóricos, emplaza la misma al *asesoramiento in situ* sobre problemas y cuestiones reales,

una alternativa cada vez más presente en la literatura sobre formación docente (Feixas et al., 2013; Margalef & Álvarez, 2005; Mauri et al., 2013; Southwell & Morgan, 2009).

Este mismo esquema de preferencias se reproduce parcialmente en el caso de la *metaevaluación*, tercera alternativa formativa sugerida por la muestra¹⁰⁰. Una parte de las/os docentes propone el desarrollo de experiencias de autorrevisión sobre los dispositivos evaluativos aplicados, mientras que el porcentaje de quienes apoyan la participación experta en dicho proceso resulta mucho menor. En la literatura específica también se reconoce la metaevaluación como una contribución relevante al desarrollo de la competencia evaluativa, quedando abierto el debate sobre los agentes que deben intervenir en su desarrollo. La metaevaluación informal, dependiente del propio docente evaluador (Stake, 2006), puede resultar parcial e imprecisa, por lo que determinadas organizaciones apuestan por la revisión experta y externa (American Evaluation Association- AEA, 2004; Quality Assurance Agency- QAA, 2014). Entre ambos extremos surgen opciones intermedias como la colaboración de colegas o el asesoramiento auxiliar de agentes expertos (Capllonch et al., 2009; Monereo & Castelló, 2009a), pero éstas no han sido señaladas en las respuestas recibidas.

El análisis de las modalidades formativas propuestas dibuja un planteamiento próximo a las comunidades profesionales de aprendizaje (Shulman, 2000) o a la idea de la *scholarship of teaching and learning* (Andresen, 2000; Boyer, 1990; Chocarro et al., 2013; Kreber, 2005), aunque en este caso con menor insistencia en el rigor y la transferencia. Centrado específicamente en el terreno de la evaluación, el profesorado se muestra partidario de someter a revisión su actuación, intercambiar experiencias y construir nuevos saberes y aprendizajes, pero a pequeña escala, sin una referencia expresa a la comunidad científica y la sistematicidad de la investigación propia de la misma.

Consideradas en su conjunto, las sugerencias recopiladas ofrecen respuesta tentativa a una pregunta compleja y delicada subrayada, entre otros, por Stiggins (1998) o Monereo y Castelló (2009a): cuando se trata de la capacitación evaluativa docente, ¿quién debe desempeñar el rol de formador? El profesorado consultado apuesta por alternativas evaluativas colegiadas, basadas en el intercambio y discusión entre iguales y, circunstancialmente, asesoramiento externo. Estas propuestas rompen con los modelos

¹⁰⁰ Considerando que la participación del profesorado en la investigación exige un ejercicio de autoevaluación de la competencia evaluativa, muy próximo a lo que supondría una metaevaluación informal en la práctica, no debe desestimarse en este resultado el posible efecto de lo que Manzano y Braña (2005) denominan la transformación de la muestra. Dicho efecto obliga a asumir que la sugerencia de la metaevaluación como modalidad formativa, podría estar motivada por la propia participación del profesorado en la investigación, dando lugar a un cierto sesgo.

formativos tradicionales, jerárquicos y lineales, y se inclinan a la constitución de *equipos generadores de formación*, más que a la designación de agentes formativos especializados (Mateo, 2012). La referencia persistente a las inquietudes reales del profesorado como núcleo principal atención deja entrever que, en esta noción del equipo, se apuesta por la horizontalidad de la relación entre agentes formadores y formados, incluso en los casos en los que se cuente con participación experta.

Los **temas** son el aspecto al que mayor atención han prestado las/os docentes a la hora de describir sus necesidades formativas con respecto a la evaluación del aprendizaje (objetivo H2). Los núcleos señalados, guardan una correspondencia razonable con las principales inquietudes señaladas desde la literatura y las organizaciones especializadas, y son coherentes con las deficiencias y dificultades puestas de manifiesto en el análisis de los objetivos anteriores. También en este caso, habría que valorar el impacto potencial de la propia investigación en las respuestas ofrecidas por la muestra.

Aunque el abanico de temas señalados es amplio y variado, la frecuencia de aparición de las diversas aportaciones permite entrever diferentes niveles de prioridad. Si la sobrecarga era apuntada por el profesorado como una de las mayores dificultades para ejercer sus funciones evaluativas (objetivo E3), la *evaluación viable o sostenible en grupos numerosos* representa una de las mayores inquietudes puestas de manifiesto. Bajo ese denominador común, las/os docentes demandan formación en herramientas, métodos y procedimientos que faciliten la evaluación continua y el feedback en tales circunstancias. Dicha necesidad también se encuentra ampliamente señalada desde la literatura (Arénaga, 2013; Ballantyne et al., 2002; Brown, 2003a; Hounsell, 2007; Liu & Carless, 2006) y representa uno de los ejes de mejora que sugieren el AR@HKU (2009a) y el CSHE (2002) para la investigación y la formación evaluativa.

A pesar de representar el tema más reiterado en la formación permanente realizada (objetivo H1), la *evaluación de competencias* figura entre los núcleos formativos sugeridos. El profesorado persiste en señalarlo entre sus necesidades, en coherencia con las dificultades puestas de manifiesto en la autoevaluación de su competencia evaluativa (objetivo E3). La actualidad del mismo resulta incontestable en vista del gran número de publicaciones, acciones formativas y jornadas que lo sitúan como su núcleo central.

En línea con las orientaciones que proporcionan Verdejo (2008), McMillan (2000) o Trillo (2005), las/os participantes reconocen precisar *formación básica en evaluación* para el desarrollo de su competencia evaluativa. Esta demanda coincide con la necesidad de preparación pedagógica reivindicada por Arénaga (2013) y Mas (2011), quien argumenta que la cualificación disciplinar queda debidamente resuelta en la formación previa y los

estudios de doctorado. También encuentra respaldo en otros resultados que ya han sido citados en la discusión de objetivos anteriores. El más relevante es la constatación de que el profesorado con formación previa en Pedagogía, Psicología y Psicopedagogía tiende a un mayor desempeño práctico en los factores asociados a su competencia evaluativa, tanto desde su propia perspectiva (objetivo E2), como desde la del alumnado (objetivo F2).

A juzgar por su presencia explícita en el listado de temas sugeridos, la *planificación* y los *principios técnicos y éticos* que deben regir el desarrollo de la evaluación deben ocupar un espacio destacado en esta preparación básica. En el caso concreto de los principios técnicos, la demanda podría estar motivada por la escasa Adecuación y precisión que atribuye el profesorado a sus dispositivos evaluativos (objetivo E1). Esta insuficiente capacitación técnica en procedimientos de evaluación y medida y la necesidad de reforzarla ya ha sido puesta de manifiesto en trabajos previos, tanto para el contexto escolar (Marso & Piegge, 1993/2012; Wise & Lukin, 1993/2012), como para el universitario (Arribas, 2012; Erwin, 2003).

A medio camino entre la evaluación de competencias y la formación básica en evaluación, las sugerencias recopiladas recogen de forma expresa el trabajo sobre *métodos, técnicas e instrumentos*. Aunque también figura entre los temas tratados en la formación permanente realizada, la muestra docente demanda apoyo formativo para descubrir o profundizar en nuevos recursos aplicables a la evaluación, necesidad que también constatan V. Álvarez et al. (2011). Sin embargo, en la literatura específica abundan los testimonios que señalan la insuficiencia de una capacitación meramente instrumental (Bordas & Cabrera, 2001; Brown, 2003b; Cano & Ion, 2012; Knight, 2005; Ramsden, 2003; Verdejo, 2008). El cambio de concepciones evaluativas requiere un proceso profundo y complejo que no parece probable sólo por el descubrimiento de técnicas y herramientas nuevas. Obviar tal consideración y limitar la formación evaluativa a la vertiente instrumental supondría incurrir en lo que Pozo (2001) califica como una práctica formativa ciega; éstas se caracterizan por la presunción de que la incorporación de un nuevo elemento teórico (instrumentos y técnicas evaluativas en este caso) basta para alterar la práctica.

Dentro de este tema, merecen una especial mención las *herramientas y recursos vinculados a las tecnologías de la información y la comunicación*. Su uso potencial en el proceso evaluativo representa una de las principales líneas de avance en el campo profesional de la evaluación del aprendizaje según se reconoce a nivel nacional (Bezanilla et al., 2014; Guenaga et al., 2013; Olmos & Rodríguez, 2010; G. Rodríguez & Ibarra, 2011, 2012) e internacional (AR@HKU, 2009a; Center for Research on Evaluation, Standards & Student Testing- CRESST, 2012; Murphy, 2006). Junto a la proliferación de recursos y herramientas en línea para el desempeño de las funciones y tareas evaluativas (simulación

de situaciones reales, programas generadores de rúbricas, aplicaciones para el diseño de pruebas de evaluación on-line, plataformas, etc.), así lo evidencia también la publicación reciente de varias tesis doctorales que analizan en profundidad el tema (Bassani, 2009; Olmos, 2008; M. M. Sánchez, 2010). En correspondencia con dicha tendencia, parte de la formación permanente realizada por la muestra versa sobre este tipo de usos y aplicaciones (objetivo H1), sin embargo su presencia es escasa entre las demandas. Con la información recabada no es posible determinar si ello se debe a la falta de interés y/o necesidad percibida o a una interpretación más amplia de estas herramientas como recursos de evaluación, lo que las situaría en el núcleo de los procedimientos y herramientas.

MARCO CONCLUSIVO

CAPÍTULO 10.

CONCLUSIONES DE LA INVESTIGACIÓN

En respuesta al planteamiento general dibujado en el capítulo introductorio, el presente trabajo ha enfocado su atención en dos preguntas centrales de investigación, respondidas en distintos apartados (ver tabla 162). La primera de ellas hace referencia a los rasgos que teóricamente deberían acompañar a la competencia evaluativa docente y ha sido abordada a lo largo del marco teórico. Efectuada una primera tarea de definición y conceptualización (*objetivo B*), su delimitación ha partido del establecimiento de una teoría de la evaluación del aprendizaje de referencia (modelo ITC), fruto de la revisión y síntesis de referentes y antecedentes relevantes (*objetivo A*). Unida a ella, se ha presentado una propuesta de sistematización de la competencia evaluativa docente, basada en la delimitación de dimensiones, indicadores y descriptores para su ejercicio (*objetivo C*). Concebido con fines pedagógicos de formación y evaluación, el modelo resultante ha sido construido considerando las conclusiones comunicadas en las contribuciones previas al tema, así como los resultados de un trabajo de campo sobre el contenido evaluativo en las escalas institucionales de evaluación de la calidad docente (*objetivo D*). La segunda pregunta de investigación ahonda en el contraste de dicho modelo en la práctica y ha sido abordada por medio del marco empírico. Contando con la perspectiva de estudiantes y profesorado, su desarrollo ha conducido a la revisión de los indicadores de la competencia evaluativa docente mejor y peor solventados en la práctica (*objetivos E, F, G*). Por su protagonismo entre los factores condicionantes y las sugerencias de mejora reportadas en la investigación y la literatura, este estudio también se ha detenido en el análisis de la formación realizada y demandada por el profesorado para su capacitación evaluativa (*objetivo H*).

Tabla 162

Repuesta articulada a las preguntas y objetivos de la investigación

PREGUNTA	OBJETIVO	RESPUESTA ELABORADA	CAPÍTULO	
COMPETENCIA EVALUATIVA EN TEORÍA: RASGOS DESEABLES	OBJETIVO A	Modelo Integrado Teórico Complejo (ITC).	IV	VI
	OBJETIVO B	Conceptualización, glosario, tipología.	II	
	OBJETIVO C	Modelo competencia evaluativa docente.	V	
	OBJETIVO D	Análisis de tópicos evaluativos en escalas institucionales de evaluación de la calidad docente.	V	
COMPETENCIA EVALUATIVA EN LA PRÁCTICA: RASGOS REALES	OBJETIVO E	MARCO EMPÍRICO. Estudio de la competencia evaluativa docente desde la perspectiva de profesorado y estudiantes de los grados asociados a la formación básica de maestras/os de Educación Infantil y Primaria.	VII, VIII	IX
	OBJETIVO F			
	OBJETIVO G			
	OBJETIVO H			

Este capítulo final queda reservado a la presentación de las principales conclusiones a las que ha dado lugar el recorrido realizado. Su estructura prevé cuatro apartados para atender a otros tantos focos de reflexión:

- Partiendo de la discusión pormenorizada de resultados efectuada al concluir el marco teórico y empírico de la investigación (capítulos sexto y noveno, respectivamente), el primer apartado presenta las conclusiones generales de la investigación.
- El segundo apartado revisa las principales contribuciones autopercebidas en el trabajo efectuado.
- El apartado tercero se detiene en las dificultades halladas en el transcurso de la investigación y en las principales limitaciones reconocidas en la misma. Valorada esta primera experiencia, también incluye una serie de sugerencias de mejora para su superación en futuras réplicas o profundizaciones.
- El capítulo concluye con la propuesta de una serie de líneas de investigación para seguir profundizando en la competencia evaluativa docente y contribuir a su futuro desarrollo.

1. Conclusiones: claves centrales de la discusión y sugerencias de mejora

Realizada la discusión específica de las dos preguntas centrales de la investigación y los propósitos que las responden, a continuación se exponen las ideas y aspectos clave en los que concluye su valoración conjunta. La indagación teórica y empírica se concreta en 12 conclusiones que sintetizan lo más relevante del análisis anticipado en los capítulos sexto y noveno:

1. La evaluación del aprendizaje en la universidad representa un foco de atención clásico en la investigación educativa. Abordado desde diversas perspectivas, el estudio de la praxis evaluativa no sólo preserva su interés, sino que adopta un protagonismo y una efervescencia creciente, que se hace visible en el número de publicaciones, investigaciones y actividades de formación y difusión sobre el tema. La búsqueda de modelos didácticos más acordes al planteamiento formativo del EEES ha supuesto un estímulo importante en su proliferación; la evaluación ha encarnado uno de los cambios pedagógicos más relevantes de la primera fase de convergencia (1998-2010) y permanece entre los retos de mejora establecidos para la fase de revisión y consolidación (2010-2020).
2. En el contexto específico de la formación básica de maestras/os la investigación evaluativa no es tan frecuente ni abundante. A pesar del doble papel que ocupa como competencia específica docente y como resultado de aprendizaje discente, no son muchos los trabajos que se han interesado por las prácticas evaluativas que tienen lugar en las facultades de educación y escuelas de magisterio. En los últimos años esta laguna está siendo cubierta gracias al impulso que está ejerciendo personal docente e investigador de varias universidades del estado¹⁰¹. En esta revitalización destacan especialmente las contribuciones efectuadas desde el área específica de la Educación Física, mientras que el grado preparatorio para maestras/os de Educación Infantil resulta el menos representado. Aunque realizan aportaciones descriptivas relevantes, estos estudios no se detienen a sistematizar la evaluación del aprendizaje, ni la competencia evaluativa del profesorado al cargo.

¹⁰¹ Universidad de Valladolid, Universidad de Murcia, Universidad Católica San Antonio de Murcia, Universidad de Castilla-La Mancha y Universidad de León principalmente, según los trabajos reseñados entre los antecedentes.

3. Las conclusiones que revelan los diagnósticos de la práctica evaluativa efectuados en la formación básica de maestras/os no difieren de las presentadas para el resto de titulaciones. Por su adscripción natural al ámbito educativo, en ella cabría presumir una mayor sensibilidad y orientación pedagógica; sin embargo, la evaluación se presenta como área de mejora y foco de insatisfacción, igual que en otras titulaciones. Esta visión resulta especialmente evidente entre el alumnado, que concede a la evaluación mayor relevancia que el sector docente y advierte un menor cumplimiento de los principios de calidad asociados a ella en el modelo ITC.
4. Tanto para estudiantes como para profesoras/es, la regulación del aprendizaje por medio del feedback representa el más importante de los factores asociados a la evaluación y su línea de mejora prioritaria. Aunque el alumnado manifieste una postura más crítica, ambos sectores coinciden en señalar que la comunicación de los resultados de aprendizaje y los procesos de retroalimentación y proacción se desarrollan con una frecuencia, profundidad e individualización insuficiente. No contamos con datos precisos sobre el poder estimulante del feedback o su carácter dialógico, pero el incumplimiento de indicadores más básicos sugiere su presencia reducida en la práctica. El profesorado señala el número elevado de estudiantes y a la sobrecarga de tareas como causas fundamentales del insuficiente cumplimiento de los rasgos asociados al feedback de calidad en el modelo ITC. El alumnado no repara en las causas, pero coincide en la necesidad de una regulación más fluida, profunda y personalizada de su aprendizaje, que le proporcione pautas de proacción para su mejora.
5. La necesidad de una mayor regulación u orientación se manifiesta indirectamente en otras facetas de la praxis evaluativa de reconocida importancia para ambos sectores. Al valorar la continuidad de la evaluación a lo largo del proceso formativo y la explicitación de las condiciones y criterios en los que se fundamenta, estudiantes y profesorado describen un nivel de cumplimiento que podría caracterizarse como literal. Los datos recabados en ésta y otras investigaciones recientes sugieren la superación general del clásico patrón de la evaluación final y opaca, aunque la continuidad y la transparencia siguen representando áreas de mejora desde una perspectiva más pedagógica:
 - a. Estudiantes y profesorado reconocen la realización de tareas evaluativas en las distintas fases del ciclo formativo, pero la intensidad de las mismas no está graduada ni coordinada de acuerdo con las exigencias del proceso de construcción del aprendizaje. La evaluación inicial sigue siendo infrecuente,

mientras que en el tramo final tiende a producirse una sobrecarga, especialmente acusada por el alumnado.

- b. En cuanto a la claridad de los criterios y condiciones evaluativas, su explicitación se lleva a cabo al comienzo del proceso formativo según docentes y discentes. Partiendo de ese nivel informativo básico, se detecta la falta de un trabajo orientado expresamente a la clarificación e interiorización de los criterios expuestos, de forma que sean unívocos y se hagan presentes a lo largo de todo el proceso. Aunque se producen ciertas discrepancias entre ambas fuentes al valorar este segundo nivel de explicitación, en ninguno de los casos puede darse por satisfecho: las actividades para el recuerdo y la clarificación o ejemplificación parecen poco extendidas y no puede hablarse del uso generalizado de rúbricas o listas de control a tal efecto.
6. La retroalimentación y la profundización en la vertiente más formativa de la continuidad y la claridad evaluativa representan áreas de mejora sentidas y reconocidas. Frente a ellas, existen otros espacios de la evaluación del aprendizaje que, a pesar de probarse insuficientemente cubiertos en la práctica, no figuran entre las necesidades apuntadas por estudiantes y profesorado. Tal es el caso de la atención a la diversidad y de la participación discente que, con independencia de su recomendación generalizada desde la perspectiva teórica, representan asignaturas pendientes en la praxis evaluativa. Así se ha evidenciado en el presente trabajo y en otros estudios similares:
- a. Para alumnado y profesorado la participación discente en el proceso de evaluación es el menos relevante de los factores considerados y, junto a la Metaevaluación, el menos llevado a cabo. Tomando distancia de las orientaciones teóricas detalladas en el modelo ITC, el análisis de la praxis evaluativa revela la preeminencia del rol docente en las diferentes fases del proceso; estudiantes y profesoras/es coinciden en señalar que el diseño evaluativo se realiza sin intervención discente, igual que las escasas experiencias de metaevaluación que se desarrollan. La participación del alumnado se reduce a experiencias puntuales de autoevaluación y coevaluación, en cuyo desarrollo parece tener mayor peso la viabilidad práctica del sistema evaluativo que su aprovechamiento formativo. A pesar de que este nivel de realización resulte insuficiente desde la perspectiva teórica, para profesorado y estudiantes la falta de participación no constituye

una de las prioridades a atender en la mejora de la evaluación. En otras palabras, no representa un área de mejora *sentida*.

- b. Frente a la evaluación participativa, que representa un área de mejora *no sentida*, la atención a la diversidad parece suscitar un cierto cuestionamiento del principio teórico que le precede. Su presencia en la evaluación representa un tema polémico y controvertido desde el punto de vista de expertas/os, docentes y discentes. La literatura específica defiende la aproximación a dispositivos evaluativos no estereotipados y sensibles con la diferencia. Sin embargo, los testimonios que se desprenden del estudio empírico sugieren que la justicia evaluativa se interpreta en términos de igualdad de condiciones y que las adecuaciones se consideran un acto de excepcionalidad o síntoma de agravio. Las directrices institucionales y la normativa evaluativa así lo predisponen, haciendo énfasis en la uniformidad y estableciendo situaciones de excepcionalidad justificables. En este contexto se constata un acuerdo general que sitúa la praxis evaluativa actual más próxima al polo de la uniformidad y la estereotipia que al de la diversificación; sin embargo, a la hora de determinar si ese debe ser o no su lugar, el acuerdo se diluye. Desde esta perspectiva, presumimos en la adecuación o diversificación evaluativa un área de mejora *no admitida*.
7. Aunque la situación descrita reporte un avance reseñable en comparación con los estudios efectuados en las dos últimas décadas, la evaluación del aprendizaje sigue representando un importante foco de atención para la mejora de la enseñanza universitaria. Así se evidencia en contraste con las pautas de deseabilidad establecidas en el modelo ITC.
- a. Considerados conjuntamente el punto de vista docente y discente, la planificación de los sistemas evaluativos destaca como el factor mejor cubierto en la práctica. Su diseño preserva la correcta alineación con objetivos formativos y metodología, y hace de la evaluación un proceso continuo que no se limita a la acreditación de resultados por medio de exámenes. Según ambas fuentes, se basa en criterios y condiciones claramente definidos y su puesta en práctica se desarrolla de forma coherente; los ajustes en los sistemas evaluativos no son frecuentes, responden a situaciones excepcionales y, por lo general, se comunican y justifican debidamente.

- b. Sin embargo, docentes y discentes acusan síntomas de un ejercicio insuficiente de la función pedagógica y reguladora en los dispositivos evaluativos, ya sea por no favorecer la integración de los criterios en los que se basan, por una graduación poco sensible de las tareas que incorporan o por deficiencias en el proceso de retroalimentación de resultados. Como ya ha sido expuesto, el proceso se desarrolla apenas sin intervención discente y de acuerdo a pautas y condiciones uniformes que no atienden a necesidades o situaciones particulares. La metaevaluación o revisión de la evaluación sigue presentándose como una práctica muy poco habitual, a pesar del beneficio potencial que se le reconoce para la mejora del proceso evaluativo, del aprendizaje y de la enseñanza.
8. Por las repercusiones adicionales que comporta, este área de mejora resulta especialmente prioritaria en las titulaciones ligadas a la formación básica de maestras/os y, por tanto, de futuras/os evaluadoras/es. Las concepciones evaluativas determinan el modo de ejercer la evaluación y se desarrollan a partir de experiencias directas que favorecen su integración y posterior reproducción. En consecuencia, la exposición reiterada a prácticas y modelos de evaluación disfuncionales incrementa su probabilidad de reproducción y perpetuación. Conectadas con los niveles inferiores del sistema educativo, en estas titulaciones las contradicciones entre la teoría y la práctica evaluativa resultan especialmente alarmantes y exigen de una mayor atención y cautela. Las orientaciones teóricas que se ofrezcan al alumnado para el ejercicio de la evaluación en el futuro próximo deben ser vivenciadas y experimentadas en primera persona en el presente, no sólo ser leídas o escuchadas.
9. Los resultados puestos de manifiesto sugieren la formación docente específica sobre evaluación como principal línea de intervención para la mejora de la competencia evaluativa. Aunque su incidencia práctica resulta minoritaria en los dos últimos años, la necesidad de capacitación evaluativa representa una afirmación ampliamente generalizada en la literatura y entre profesorado consultado. Las diferencias detectadas en la praxis evaluativa en función de la titulación docente de acceso sugieren que dicha formación debería adoptar una orientación prioritariamente psicopedagógica. Tituladas/os en Psicología, Pedagogía y Psicopedagogía refieren mayores niveles de realización en los factores e indicadores asociados a la competencia evaluativa; el alumnado de primer curso y del módulo de formación básica (donde mayor concentración de profesorado con

estas titulaciones se produce) también coincide en observar un mejor cumplimiento. Una posible causa de esta diferencia pueden ser las dificultades generales asociadas a la falta de formación específica en el acceso a la docencia universitaria. Especialmente evidente en lo que respecta a la capacitación evaluativa, la investigación y la literatura sugiere que esta carencia predispone al profesorado a la reproducción inconsciente de las deficiencias propias de los clásicos patrones de evaluación. En el contexto formativo acotado, este efecto podría dar lugar a su extensión en próximas generaciones de maestras/os.

10. La efectividad potencial de la formación específica no depende tanto de la ampliación de su oferta, como de la satisfacción de una serie de requisitos y condiciones que favorezcan el impacto sobre la identidad profesional docente y las concepciones evaluativas sostenidas. Las necesidades y preferencias identificadas en el estudio teórico y empírico sugieren una determinada orientación para dicha formación. Las conclusiones obtenidas sirven para dar respuesta a algunas de las inquietudes de las que da cuenta la literatura, ya sea para debatir el planteamiento de la formación docente general (Margalef & Álvarez, 2005; Montero, 2006; Zabalza, 2002) o expresamente la capacitación evaluativa (Stiggins, 1998):

- a. *¿A quién debe dirigirse la formación específica en evaluación?* Concebida como núcleo de interés institucional, sus destinatarias/os potenciales serían todas aquellas personas que vayan a asumir funciones docentes en la formación básica de maestras/os, independientemente del tipo de contrato y vinculación profesional. La oferta debería orientarse especialmente a las/os profesionales cuya formación previa no ha procurado una preparación y sensibilización psicopedagógica para el ejercicio de la enseñanza.
- b. *¿En qué momento debe realizarse la formación específica en evaluación?* La mejora de la competencia evaluativa puede favorecerse por medio de programas de formación inicial y permanente. Las condiciones que actualmente regulan el acceso a la docencia universitaria no contemplan ningún requisito formativo específico vinculado a la enseñanza, salvo imposición expresa de la propia institución contratante. En tanto que no se modifiquen tales condiciones, las acciones e iniciativas orientadas al desarrollo de la competencia evaluativa deberían integrarse preferentemente dentro de los programas de formación permanente para el profesorado. No obstante, a un nivel superior debería considerarse la instalación de

programas de acceso o iniciación a la docencia¹⁰², que mitiguen las repercusiones de los modelos de preparación consecutiva en la identidad profesional y el ejercicio docente.

- c. *¿Qué carácter debe tener la formación específica en evaluación?* En el marco general de acceso a la docencia dibujado, la formación permanente en evaluación debería llevarse a cabo de forma voluntaria, de modo que pueda tener lugar la reflexión y retrospección que exige el cambio de concepciones evaluativas. Esta alternativa exige admitir los riesgos del voluntarismo, más aun considerando la amplia distancia que se evidencia entre la necesidad de formación evaluativa comunicada por el profesorado y su realización efectiva. Para contrarrestarlos, esta preparación debería ser reconocida y valorada en procesos de acreditación, promoción, etc.
- d. *¿Cuáles son los contenidos a los que debe atender la formación evaluativa?* La formación evaluativa debería promover la integración y el desarrollo práctico de los criterios e indicadores asociados a la calidad de la evaluación en el modelo ITC. Partiendo de este marco de referencia común, entendemos que la formación evaluativa debería trabajar a dos niveles consecutivos, ligados a inquietudes con distinta prioridad. En un primer momento deberían atenderse las necesidades evaluativas reconocidas y sentidas por el profesorado, para abordar posteriormente aquellos otros principios de calidad que no parecen asumidos entre los referentes comunes de deseabilidad. En ambos casos, debería favorecerse una aproximación fundamentalmente psicopedagógica al fenómeno evaluativo.

Dentro de las necesidades directamente señaladas por el profesorado destaca el tema de la evaluación de competencias y su puesta en práctica en grupos numerosos. Para su abordaje, la formación debería plantearse desde un marco de reflexión amplio, que no corresponda literalmente al protagonismo que ocupan los instrumentos de evaluación en las demandas docentes. Sin desoírlos, la profundización sobre técnicas y herramientas debería ponerse al servicio de otras áreas de mejora detectadas en la evaluación, no como un fin en sí misma. Por su significación y relevancia, la regulación del aprendizaje por medio del feedback constituiría la más

¹⁰² Se desarrollan desde hace años en otros países de la Unión Europea (Alemania, Francia, Suiza, Holanda, Reino Unido, Noruega, etc.), Canadá o Estados Unidos (Perales, Sánchez, & Chiva, 2002; Torra et al., 2013).

relevante de todas ellas, seguida del refuerzo de la continuidad y la transparencia de la evaluación desde una perspectiva más pedagógica. Este primer nivel de formación también debería atender a la necesidad de clarificación de conceptos y principios evaluativos que, aunque manida, parece permanecer insatisfecha.

Sin perjuicio de lo anterior, en un segundo nivel la formación debería orientarse a suscitar la reflexión sobre rasgos evaluativos deseables desde la perspectiva teórica que, por el momento, no parecen gozar del respaldo de estudiantes y profesorado. Entre ellos, parece especialmente urgente la profundización en los límites y las posibilidades de dos focos destacados de atención: la participación discente en la evaluación del aprendizaje y la diversificación o atención a la diversidad en la misma.

- e. *¿Cuáles deben ser las modalidades preferentes para la formación evaluativa?* Las modalidades formativas por las que manifiesta predilección el profesorado reúnen los principales requisitos señalados para el cambio en las concepciones evaluativas. En principio, el clásico formato expositivo, basado en la presentación teórica unilateral por parte de expertas/os, debería ser secundario en la oferta principal. En su lugar, se debería apostar por formatos basados en el intercambio, la discusión y la difusión horizontal de experiencias, en un espacio para el encuentro y la participación directa del profesorado. Abundan las referencias a actividades y prácticas formativas contextualizadas, basadas en problemas y situaciones evaluativas reales, en cuyo tratamiento también tiene cabida el asesoramiento experto. Desarrollada bajo dirección propia o de forma colaborativa, la metaevaluación también debería contemplarse dentro del repertorio de alternativas formativas para el desarrollo de la competencia evaluativa. Tanto en la revisión, como en actividades de profundización, discusión o intercambio, el trabajo conjunto del profesorado de la misma institución, favorecería además procesos de coordinación en materia evaluativa y, con ello, la respuesta a otra de sus limitaciones clásicas.
- f. *¿Quién debe desempeñar el rol de agente formador en estas actividades?* Las modalidades formativas propuestas diluyen, en cierto modo, la barrera que separa a los agentes formadores y formados. Basadas en el intercambio y la discusión horizontal, las alternativas sugeridas dan lugar a toda una corriente de formación interna, que implica a docentes de la propia institución formativa en la revisión, resolución o discusión de situaciones evaluativas

reales. Esta participación colegiada también podría extenderse a profesorado de otros centros o universidades por medio de redes profesionales o encuentros presenciales periódicos. Para contrarrestar el riesgo de imprecisión o la falta de sistematización, cabría complementar estas alternativas con la intervención de agentes expertos especializados, pero en calidad de asesores, no como ponentes. En este contexto, la labor de los centros de formación del profesorado sería propiciar la creación de núcleos de formación, más que desempeñar el rol formador directamente. Entre sus funciones también recaería la creación de los programas de formación a los que se alude a continuación.

- g. *¿Cuánta formación evaluativa es necesaria?* Lejos de pretender una cuantificación precisa, baste con aclarar que, hasta la fecha, la formación docente en materia evaluativa resulta insuficiente. Frente a esta realidad, se hace cada vez más evidente la necesidad de procurar al profesorado la preparación que requiere este complejo e importante proceso. Para ello, se sugiere la creación de programas de formación docente amplios, que vayan más allá de un ciclo de acciones puntuales. Aunque no exista un marco regulador que lo requiera, la complejidad de la profesión docente hace conveniente la previsión de una variedad de actividades de formación inicial y permanente para profundizar en la evaluación y el resto de competencias específicas ligadas a la enseñanza. Frente a las clásicas acciones formativas, puntuales y expositivas, conviene la previsión de secuencias de formación suficientemente extensas y dilatadas para posibilitar la puesta en práctica y la proacción sobre los elementos trabajados. Éstas deberían generalizar la imagen de la evaluación del aprendizaje como un objeto de análisis complejo, que no está sujeto a aproximaciones e interpretaciones únicas e incuestionables.

11. El desarrollo de ese amplio abanico de acciones formativas requiere el establecimiento de un marco de referencia claro, que defina indicadores de calidad en la competencia evaluativa y ofrezca un soporte para orientar las sugerencias y aproximaciones sucesivas de mejora. Basada en la teoría de la evaluación detallada en el modelo ITC, la propuesta de sistematización configurada pretende servir a dicha necesidad, desglosando la competencia evaluativa en dimensiones, indicadores y descriptores de logro. Empleada como soporte para la formación y/o la metaevaluación, su uso puede tener lugar en modalidades autodirigidas o compartidas y contando o no con asesoramiento externo. Asumida como referencia

compartida, también propicia un marco común para establecer procesos de coordinación e intercambio. La complejidad y amplitud de la propuesta, exige un trabajo de aproximación gradual y progresivo, que intercale la revisión y la intervención de acuerdo a las prioridades y posibilidades que se identifiquen en cada caso.

12. Por último, el desarrollo de la competencia evaluativa docente no debe desligarse de la revisión de las condiciones estructurales que rodean la evaluación, ejerciendo un poderoso influjo en su desempeño. Entre ellas, las más urgentes parecen la reducción del elevado número de estudiantes y la revisión y actualización de la normativa evaluativa, ambas necesarias para responder a los retos evaluativos que plantea la filosofía evaluativa del EEES. Estas mejoras deberían tener lugar en un marco de reflexión más amplio que atienda al reconocimiento y fortalecimiento de la identidad profesional docente.

A pesar del avance constatado con respecto a los estudios desarrollados en las últimas décadas, las áreas de mejora atribuidas a la evaluación del aprendizaje siguen poniendo de manifiesto la necesidad de reforzar su carácter formativo y su función pedagógica (mejora de la retroalimentación y proacción, avance hacia la continuidad sensible, interiorización de criterios y condiciones, etc.). Así lo reflejó también I. Álvarez (2008b), efectuada una revisión de las conclusiones comunicadas en publicaciones científicas sobre la evaluación del aprendizaje en la universidad entre 1996 y 2007. Aún de forma más reciente, las recomendaciones reportadas en estudios próximos (Gutiérrez et al., 2013; Hortigüela et al., 2015; Villa et al., 2013) persisten en la misma apreciación.

El cambio en las prácticas evaluativas no representa una empresa fácil a pesar del acuerdo general que suscita su conveniencia. De acuerdo con Serafini (2000), el reto consiste en consolidar el cambio en el paradigma de la evaluación, lo que afecta a la administración educativa, a los programas de desarrollo profesional y a los modelos de formación docente, no sólo al profesorado en el que se encarna la competencia evaluativa. Campbell (1993/2012) va más allá y afirma que

en la medida en la que la evaluación de los estudiantes se diseñe en respuesta a creencias obsoletas sobre el conocimiento, la mente, el aprendizaje y las organizaciones humanas, el rol y las responsabilidades del profesorado en la conducción de dicha evaluación, sus propósitos al hacerlo, y la formación que reciban para satisfacer dichas necesidades, serán igualmente obsoletas. (p. 97).

Por representar un enclave estratégico con respecto a la estructura general del sistema educativo, nuestra contribución selecciona la formación básica de maestras/os de Educación Infantil y Primaria como espacio de intervención prioritaria para iniciar este proceso de replanteamiento y avance. Esta acotación es consistente con las recomendaciones de Schommer et al. (2012), V. M. López y Palacios (2012) o Palacios y López (2013), que reconocen que la mejora de las prácticas evaluativas resulta especialmente urgente en dichas titulaciones.

2. Contribuciones de la investigación

Valorada con cierta distancia, encontramos que las principales contribuciones de la investigación efectuada se resumen en torno a tres núcleos centrales sobre los que se profundiza a continuación.

La delimitación de una teoría evaluativa de referencia y la propuesta de un modelo de competencia docente para su materialización práctica representa la mayor de las contribuciones del trabajo efectuado. Ante la divergencia de tendencias, creencias e interpretaciones que se producen en torno al fenómeno evaluativo, esta aportación responde a la necesidad de sistematización de la que se hacen eco diversas investigaciones (Buendía & López, 1999; Cano et al., 2010; Padilla et al., 2010; Porto, 2006).

- La teoría de la evaluación elaborada se concreta en el modelo ITC, producto de la revisión, síntesis y estructuración de distintas referencias para el desempeño de la función evaluativa. Éste propone una definición para la evaluación del aprendizaje de calidad en el contexto específico de la formación básica de maestras/os y la operativiza en la delimitación de siete subprocesos, 19 criterios y 68 indicadores. Juntos, configuran un sistema ordenado que sintetiza los aspectos clave presentes en la literatura, en los requerimientos normativos y contextuales, así como en los decálogos, modelos o estándares que previamente han intentado orientar la praxis evaluativa.
- El trabajo realizado incorpora también un modelo de competencia evaluativa docente para favorecer la transferencia práctica de los referentes de calidad establecidos en la propuesta ITC. El mismo contempla ocho dimensiones relevantes para el ejercicio

de la evaluación en la formación básica de maestras/os, delimitando para cada una de ellas indicadores y descriptores de logro que clarifican su interpretación. Frente a la gran cantidad de referencias y aportaciones que se producen en torno a la evaluación del aprendizaje universitario, esta propuesta viene a suplir la escasa atención y sistematización que proporciona la investigación y la literatura específica de la competencia evaluativa del profesorado.

- Ambas propuestas ofrecen una visión sintética y ordenada que favorece la satisfacción de dos demandas clásicas, reiteradas en diversos foros: la coordinación evaluativa (Bolarín et al., 2013; Briggs et al., 2003; Gairín, 2009; Porto, 2008, 2010; Sayós, 2010) y el entendimiento mutuo (J. M. Álvarez, 2003; Arribas et al., 2010; JCSEE, 2007; V. M. López et al., 2007). El establecimiento de un glosario y la revisión de las principales tipologías evaluativas durante su elaboración (**ANEXOS 02 y 03**) resulta de particular utilidad ante las dificultades habituales que generan la ambigüedad y la confusión terminológica.
- El modelo ITC y la propuesta de competencia evaluativa que lo complementa ofrecen un soporte para orientar la evaluación del aprendizaje en las titulaciones vinculadas a la formación básica de maestras/os. Sin embargo, representan referencias amplias y flexibles, susceptibles de ser adoptadas desde instituciones formativas diferentes y distintos estilos docentes. Los indicadores incorporados en ambas propuestas pueden ser abordados de forma selectiva o ser adecuados, en función de las necesidades, prioridades o singularidades presentes en cada caso.

La segunda contribución destacable en este marco de conclusiones es la definición de una *rúbrica o rejilla de valoración como soporte al modelo de competencia evaluativa*. En ella se detallan cinco niveles de consecución para cada uno de los indicadores de calidad formulados. El sistema de descriptores configurado ayuda a clarificar el sentido y significado de cada uno de ellos, favoreciendo el desempeño efectivo de las funciones formativa y evaluativa a las que pretende servir la propuesta. En correspondencia con los beneficios reconocidos a este tipo de herramientas (I. Álvarez & Olivera-Smith, 2014; Andrade, 2000, 2005; Panadero & Jonsson, 2013), el nivel de detalle y pormenorización alcanzado facilita su aplicación, incluso sin asesoramiento externo. En este sentido, la rúbrica elaborada representa un valor añadido en la investigación, al ofrecer un recurso para el desarrollo autónomo de la competencia evaluativa a nivel individual o colegiado:

- Empleada como herramienta para la metaevaluación, la rúbrica propuesta contribuye a superar la clásica dificultad de indefinición en cuanto a los aspectos concretos a revisar (Schwandt & Halpern, 1988). Aunque también puede ser

empleado bajo dirección externa, este recurso resulta particularmente útil para propiciar procesos de revisión interna, desarrollados individualmente o en colaboración con colegas. Por su configuración, el trabajo con la rúbrica podría representar en sí misma una experiencia formativa, exponiendo referentes de deseabilidad y contribuyendo al cuestionamiento crítico de las propias prácticas evaluativas. Así lo sugieren las observaciones de Hills y Glover (2005) sobre el proceso de validación del prestigioso *Assessment Experience Questionnaire* (AEC) de Gibbs y Simpson (2003).

- La rúbrica presentada puede ser empleada para guiar el planteamiento de sistemas de evaluación, no sólo su revisión. En este segundo uso o funcionalidad, los indicadores y descriptores desglosados en la propuesta detallarían pautas hacia las que dirigir la planificación y la puesta en práctica de los dispositivos evaluativos. Como en el caso de la metaevaluación, esta posibilidad resulta particularmente interesante para su empleo interno por parte del profesorado, ya sea individualmente o de forma colegiada. En consecuencia, ofrece un recurso para satisfacer las necesidades de formación evaluativa anteriormente referidas. Cuando se emplea de forma compartida, con expertas/os o colegas, además proporciona un soporte para articular procesos de discusión e intercambio, en línea con las recomendaciones presentadas para su desarrollo.

Por último, *el foco de atención* representa a nuestro juicio, otra de las contribuciones atribuibles al trabajo llevado a cabo. En comparación con otras investigaciones próximas, *el análisis realizado ofrece una perspectiva amplia de la evaluación del aprendizaje*. No se limita a aspectos parciales como la instrumentación, la frecuencia o el grado de participación y ofrece una visión compleja y multidimensional de la praxis evaluativa y de la competencia docente asociada.

3. Dificultades, limitaciones y sugerencias de mejora

El desarrollo del trabajo presentado ha exigido la superación de ciertas **dificultades específicas**, además de las propias y comunes a todo proceso de investigación. Entre las que mayor impacto han ocasionado destacan las siguientes:

- Frente a otros trabajos que se han centrado en aspectos parciales de la evaluación del aprendizaje, la presente investigación aboga por un enfoque amplio de la misma que abarca su planteamiento general en distintas dimensiones. Aunque esa perspectiva representa un valor añadido, su materialización entraña dificultades operativas considerables. La más relevante de todas ellas es el elevado número de indicadores a incorporar en la valoración de la competencia evaluativa, dando lugar a herramientas de recogida de información extensas y complejas.
- Como ya ha sido reseñado en distintos capítulos, la evaluación del aprendizaje en la universidad no es un tema novedoso, aunque preserva su actualidad por el interés que suscita. Por presión o impulso de los planteamientos del EEES, su efervescencia en los últimos años resulta especialmente manifiesta, favoreciendo la difusión y transferencia de numerosos estudios y publicaciones sobre el tema. En contrapartida, la evaluación del aprendizaje se convierte en un tema prácticamente inabarcable, en el que no dejan de sucederse aportaciones significativas y relevantes. Esta circunstancia ha derivado en un proceso permanente de revisión y actualización de fuentes, generando un listado de referencias muy extenso, pero siempre inconcluso. La adopción de una visión amplia en el análisis de la competencia evaluativa agrava esta dificultad al multiplicar los focos de atención y búsqueda.
- La accesibilidad a la muestra ha supuesto otra dificultad reseñable en las fases iniciales del proceso investigador. La evaluación del aprendizaje permanece a menudo en el terreno de lo privado (Zabalza, 2002), de modo que la captación de profesorado, dispuesto a evaluar su propia competencia evaluativa con intervención del alumnado, ha representado una primera barrera a superar. El acceso a estudiantes ha resultado aún más complejo por las propias exigencias internas de la investigación (organización de una sesión de recogida presencial, en el tramo final del cuatrimestre).

La influencia indirecta de tales dificultades, junto a la confluencia de toda una serie de factores de orden profesional y práctico (inexperiencia en el ámbito de la investigación, accesibilidad y recursos limitados, etc.), hace que las conclusiones y contribuciones de la investigación deban ser interpretadas dentro de unos **límites reconocidos**. El primero de ellos hace referencia a la *magnitud relativa del objeto de estudio*, en el marco general de competencias profesionales atribuibles al profesorado universitario. Aunque haya sido planteada de forma amplia, la competencia evaluativa docente se ha limitado a la evaluación del aprendizaje discente, sin considerar otros espacios de evaluación que

también incumben al profesorado (el propio desempeño docente, las redes de coordinación internas o externas o el diseño y puesta en práctica de los planes de estudio, por ejemplo). Junto a ello, es importante considerar que este nivel de evaluación se circunscribe únicamente a la docencia, el primero de los cuatro roles atribuidos al profesorado, junto a la gestión, la investigación y la transferencia. Desde esta perspectiva, admitimos con V. Álvarez et al. (1999) y Mateo (2012) la artificialidad que pueda provocar esta visión fragmentada, que no capta la labor docente en su globalidad, ni considera el balance desigual que se produce entre los cuatro espacios que aglutina.

Junto a esa artificialidad inicial, en el análisis del trabajo efectuado detectamos tres *limitaciones de carácter metodológico* detalladas a continuación. Su presentación se realiza junto a **sugerencias de mejora** derivadas de esta primera experiencia, llamadas a contrarrestar su efecto potencial en futuras réplicas o profundizaciones.

- El *elevado número de indicadores* considerado en el análisis de la competencia evaluativa docente representa una dificultad importante en la configuración de instrumentos para la recogida de información. Tanto en la versión docente como en la discente, la valoración de ocho dimensiones iniciales sobre el objeto de estudio predispone al diseño de instrumentos excesivamente extensos, con los consiguientes riesgos de desatención y fatiga en el proceso de respuesta. El empleo de rúbricas repercute positivamente en el valor formativo del modelo, pero incrementa la probabilidad de sobrecarga al incorporar cinco niveles o descriptores de cumplimiento por cada indicador considerado. En virtud del carácter descriptivo y exploratorio de la presente investigación, en la fase de diseño se adoptaron una serie de decisiones prácticas para mitigar ese efecto indeseado, aunque éstas repercuten en una serie de *limitaciones metodológicas*:

- ✓ *Reducción del número de indicadores por dimensión y su repercusión en la validez como estructura interna.* Para evitar la sobrecarga mencionada, en la fase de diseño se ha minimizado en lo posible el número de indicadores por dimensión. Esta medida afecta especialmente a las dimensiones segunda (Claridad de las condiciones de la evaluación) y séptima (Participación discente en la evaluación), que quedan conformadas por un único elemento y, en consecuencia predispuestas a formar parte de otro factor más amplio. La consecuencia de esta economía de indicadores es una estructura interna razonable, aunque evidentemente mejorable desde el punto de vista psicométrico.

Para mejorar la estructura interna de las herramientas empleadas en la recogida de información debería revisarse la rúbrica de referencia, de modo que cada dimensión incorpore al menos tres elementos, el mínimo para conformar un factor en un análisis factorial exploratorio (Morales, 2013; Osborne & Costelo, 2005). Esta solución puede lograrse ampliando el número de indicadores (e incrementando con ello el riesgo de desatención y fatiga) o adecuando y reordenando la distribución inicial de indicadores y dimensiones, de forma que se satisfaga el criterio expuesto. Analizados los factores resultantes en la muestra docente y discente, nuestra propuesta de rectificación recurre a la combinación de ambas alternativas de ajuste según se presenta en el **ANEXO 20**.

- ✓ *Reducción del número de indicadores por dimensión y repercusión en la fiabilidad como consistencia interna.* Otra de las repercusiones que cabe atribuir a la reducción del número de ítems o indicadores es la escasa consistencia interna revelada para la realización del factor sexto de la herramienta docente (Puesta en práctica de la evaluación), excluido del estudio inferencial por riesgo psicométrico. Aunque en ella hayan podido incidir otros aspectos como la homogeneidad de las respuestas o la propia complejidad del objeto valorado (Morales, 2007), la modificación sugerida en el punto anterior, probablemente mejoraría la precisión de medida evidenciada, posibilitando su consideración en próximas exploraciones.
- En segundo lugar, aunque el *empleo de rúbricas o rejillas de valoración* comporte los beneficios que ya han sido reseñados a lo largo del trabajo, entrañan también ciertos riesgos y limitaciones que es preciso considerar:
 - ✓ *Definición compleja.* Orientadas a desglosar la competencia evaluativa docente en indicadores y descriptores, el proceso de elaboración de las rúbricas empleadas en ambos cuestionarios ha resultado complejo y muy costoso. La dificultad principal ha estribado en la delimitación de cinco niveles de realización en cada indicador cuestionado, superando dos retos importantes: la previsión de toda la casuística posible y la selección de descriptores que reflejen la integración de recursos cognitivos asociada a las competencias (conocimientos, procedimientos y elementos actitudinales).

La definición de una rúbrica de valoración fue una de las sugerencias derivadas de la consulta a expertas/os durante la fase de definición del modelo de competencia evaluativa docente. En respuesta a la misma, se

han realizado varias fases de discusión y revisión interna para su definición y dos aplicaciones piloto con alumnado y profesorado. En este proceso, hubiera sido conveniente realizar una segunda consulta a expertas/as, pero ésta quedó descartada en vista de las dificultades de accesibilidad y el tiempo de respuesta consumido en la primera consulta. En estas circunstancias, sugerimos que la validación de la propuesta revisada con la que concluye el presente trabajo (**ANEXO 20**) disponga de los medios oportunos para comenzar por realizar dicho contraste.

- ✓ *Dificultades de aplicación: rigidez, rudeza de medida y disparidad de interpretaciones.* Ligadas, en parte, a las dificultades que acompañan a su definición, el empleo de rúbricas como recurso para la evaluación entraña ciertas dificultades. La singularidad propia de los procesos y contextos educativos hace que puedan resultar un instrumento rígido para las/os usuarias/os, a pesar de los cinco niveles de descripción contemplados. Su administración también ha podido dar lugar a una cierta rudeza de medida (Siegel, 1975), al solicitar a las/os informantes que se posicionen en un determinado descriptor, siendo probable la combinación de matices de varios de ellos en su situación real. Por último, a pesar de que las rúbricas aporten claridad en la interpretación de aquello que se evalúa, no evitan por sí mismas la divergencia de interpretaciones (J. Hafner & Hafner, 2003; Reddy & Andrade, 2010). Este es el motivo por el que la recogida de información discente se ha llevado a cabo en sesiones grupales dirigidas, reduciendo el riesgo de ambigüedad y favoreciendo el apoyo que desde la investigación específica se recomienda brindar (Andrade, 2005; Panadero & Jonsson, 2013). Sin embargo, esta misma medida no ha podido ser adoptada en el caso del profesorado (la supervisión individual podría vulnerar su anonimato y la recogida grupal exigiría una coordinación de agendas inviable desde la perspectiva práctica).

Ante estas limitaciones, de cara a próximas réplicas o profundizaciones sugerimos mantener la aplicación guiada de las rúbricas, siempre que sea posible, y emprender dos mejoras. La primera de ellas sería complementar la administración de cuestionarios (rúbricas) con otras metodologías más cualitativas que permitan captar matices y situaciones particulares (entrevista, análisis de contenido, grupos de discusión, etc.). Aunque suponga un aumento del coste de la investigación, ello permitiría reducir la artificialidad de las respuestas y mejorar su riqueza y profundidad. La

segunda sería permitir en los instrumentos de recogida alternativas de respuesta intermedia entre los cinco niveles de realización, siempre y cuando la postura se aclare o justifique por medio de una observación.

- ✓ *Recurso más complejo desde el punto de vista psicométrico.* Sin perjuicio de los beneficios formativos y ventajas reportados para las rúbricas, la satisfacción de los requisitos psicométricos de validez y fiabilidad parece más sencilla y probable en otro tipo de instrumentos como, por ejemplo, las escalas tipo Likert. Al no estar sujetas a la delimitación de distintos niveles de logro, éstas posibilitan una mayor concentración de ítems en la consulta, favoreciendo una estructuración interna más clara y consistente.
- La última de las limitaciones a reseñar en este apartado tiene que ver con los rasgos propios del diseño de investigación no experimental o descriptivo configurado. La imposibilidad de manipulación de todas las variables implicadas provoca que *no puedan garantizarse la validez del proceso de respuesta, ni la basada en la relación con otros criterios* (AERA, APA, & NCME, 2014). La principal consecuencia de ello, es la posible incidencia de sesgos habituales como la deseabilidad social (Morales, 2006b) o la transformación de la muestra (Manzano & Braña, 2005).

Las medidas que sugiere la literatura para contrarrestar las desviaciones asociadas a dichos sesgos son relativas y en ningún caso garantizan el rigor de los estudios experimentales o cuasi-experimentales. No obstante, en próximas aplicaciones el riesgo de desviación podría disminuirse por medio de tres estrategias:

- ✓ *Ante el riesgo de deseabilidad, sensibilización.* Siguiendo el ejemplo de B. García et al. (2012) el riesgo de deseabilidad social en el colectivo docente podría mitigarse por medio de sesiones previas de sensibilización, que enfatizan la necesidad de partir de un diagnóstico realista para poder mejorar la competencia evaluativa.
- ✓ *Ante el riesgo de transformación de la muestra, dilación.* Como recomiendan Manzano y Braña (2005) el tiempo es el mejor aliado para mitigar el efecto de la propia investigación sobre la muestra. En nuestro caso, dicho efecto se hace especialmente perceptible en la sugerencia de áreas de mejora evaluativas en las preguntas finales abiertas del cuestionario. La réplica de tales cuestiones, transcurrido un cierto tiempo de la recogida de información, podría representar un medio eficaz para neutralizarlo. Otra alternativa sería fragmentar el cuestionario y administrar la parte cerrada (rúbrica) y la abierta (valoración de fortalezas y sugerencias de cambio) en sesiones diferentes.

- ✓ *Ampliación de agentes informantes.* La consulta de fuentes complementarias representa una medida para incrementar la credibilidad de las respuestas registradas y reducir el efecto de los sesgos clásicos al considerar el punto de vista de los agentes directamente involucrados. De acuerdo con Berk (2005) o P. González y Grande (1999), resulta particularmente interesante la consulta a alumnado egresado, y así se ha experimentado en investigaciones próximas (Arribas et al., 2010; Flórez, 1999; Gutiérrez et al., 2013; Hamodi & López, 2012). Aunque con un cierto riesgo de distorsión por el recuerdo de los últimos años de formación, su posición permite efectuar valoraciones libres de toda presión.

4. Líneas de investigación

Para dar cierre a la presentación de las conclusiones finales, se exponen a continuación una serie de líneas de investigación sobre las que sería interesante profundizar en el futuro. Algunas de ellas están relacionadas con las conclusiones que han sido presentadas a lo largo de este capítulo, mientras que otras hacen referencia a cuestiones relacionadas que no han podido ser abarcadas en esta aproximación inicial.

- *Adecuación del modelo para hacerlo extensible a otras titulaciones y contextos formativos.* El presente trabajo ha planteado una aproximación a la competencia evaluativa del profesorado de los grados vinculados a la formación básica de maestras/os de Educación Infantil y Primaria. Tanto el modelo ITC de evaluación como la propuesta de sistematización de la competencia docente asociada se circunscriben a dicho espacio formativo, pero sería interesante su adecuación a otras titulaciones:

- ✓ Para favorecer un mayor impacto sobre el sistema educativo, convendría trasladar el trabajo realizado a la definición de la competencia evaluativa del profesorado del *Máster de Educación*, vinculado a capacitación profesional para la enseñanza secundaria. V. M. López y Palacios (2012) señalan diferencias importantes entre la praxis evaluativa desarrollada en la formación básica de maestras/os y la que tiene lugar en ese caso; no obstante, consideramos que los referentes teóricos elaborados en el

presente trabajo resultan pertinentes y transferibles al mismo. Los matices y adecuaciones previsibles sobre el modelo ITC y la propuesta de competencia docente asociada no parecen numerosas, pero apuntan presumiblemente a la más polémica de las dimensiones evaluativas consideradas: la atención a la diversidad. Así lo sostenemos valorando una mayor heterogeneidad en el perfil discente y en las propias condiciones organizativas de la formación de postgrado (distintos grados de presencialidad y dedicación, alumnado de mayor edad con trayectorias académicas y profesionales más diversas, etc.).

- ✓ Aunque tengan lugar en la formación básica de maestras/os, la valoración de la competencia evaluativa en el contexto del *Practicum* y del *Trabajo de Fin de Grado (TFG)* parece requerir mayor adecuación. Excluidas del presente trabajo por sus particularidades organizativas y las repercusiones evaluativas parejas (implicación de diversos agentes evaluativos, desarrollo en contexto auténtico, etc.), conforman un módulo normativo e indispensable para la consecución de las competencias previstas en el perfil profesional regulado (Orden ECI/3854/2007; Orden ECI/3857/2007). En consecuencia, la evaluación del aprendizaje que tiene lugar en su desarrollo debería recibir un abordaje específico en próximas aproximaciones. La creciente difusión de contribuciones y experiencias relacionadas con el tema en redes profesionales y revistas especializadas respalda dicha necesidad en la enseñanza universitaria en general (Cid, 2012; R. Fuertes & Balaguer, 2012; Mateo, 2009a; R. Ruiz, 2011; Tejada, 2005) y, en particular, en las titulaciones que nos ocupan (Calbó, 2009; Moril, Ballester, & Martínez, 2012).
- ✓ La teoría de la evaluación y de la competencia evaluativa elaboradas en el presente trabajo también tienen sentido fuera del contexto específico de la formación de maestras/os y profesorado. Efectuadas ciertas adecuaciones sobre el enfoque y la terminología empleada, su uso como herramienta de formación y/o evaluación podría contribuir a la mejora de la calidad de la enseñanza y del aprendizaje en otras titulaciones universitarias.

- *Profundización en indicadores clave de la competencia evaluativa docente.* La racionalización del número de indicadores a considerar en cada dimensión de la competencia evaluativa ha condicionado poderosamente la profundidad del análisis efectuado. En consecuencia, consideramos oportuno seguir profundizando sobre tres indicadores clave previstos en el modelo ITC, que no han sido suficientemente desarrollados en su transferencia a la rúbrica para la valoración de la competencia

evaluativa docente. Enlazando con otra de las limitaciones advertidas en la presente investigación, en esta profundización sería conveniente recurrir a un enfoque metodológico más cualitativo, que permita indagar y profundizar en las valoraciones procedentes de estudiantes y/o profesorado:

- ✓ *El proceso de retroalimentación o feedback.* Señalado como el elemento clave para la evaluación del aprendizaje desde el punto de vista docente y discente, representa una de las principales áreas de mejora a superar en el futuro próximo. Admitido el importante papel que desempeña en el proceso de enseñanza y aprendizaje como recurso para la regulación y la autorregulación, consideramos que es preciso sumar contribuciones y avanzar en una importante línea de investigación que ya se encuentra abierta (I. Álvarez & Olivera-Smith, 2014; Cano, 2014; Evans, 2013; Jessop, 2013; Molloy & Boud, 2013; Nicol, 2013; Nicol et al., 2014; Orsmond et al., 2013; Price et al., 2013; Lu & Law, 2012). Más allá de indagar sobre la frecuencia y el formato de comunicación de los resultados de aprendizaje, ésta atiende a indicadores de calidad más exigentes como el carácter dialógico y proactivo del feedback, la intervención discente en el mismo o su poder estimulante, todos ellos contemplados en el modelo ITC. Las posibilidades que ofrecen las tecnologías de la información y la comunicación en el proceso de retroalimentación representa un foco de atención destacado en este sentido, especialmente, en el trabajo con grupos numerosos.
- ✓ *La continuidad evaluativa pedagógica y sensible.* Los resultados presentados sugieren que el tópico de la evaluación exclusivamente final se encuentra prácticamente superado. La continuidad de la evaluación y su presencia a lo largo del proceso formativo se va generalizando según arrojan las conclusiones del presente trabajo y de otras investigaciones próximas; sin embargo éstas derivan de observaciones fundamentalmente instrumentales, la mayor parte de las veces, centradas en el número de pruebas que se suceden durante el ciclo formativo para determinar la calificación.

En su lugar, consideramos oportuno indagar sobre la integración de la evaluación en el proceso formativo desde una perspectiva más pedagógica, atendiendo al empleo estratégico que se realiza de la misma para favorecer la regulación y la autorregulación en la construcción del aprendizaje discente. Esta línea complementaria de investigación analizaría no sólo el número de pruebas intercaladas, sino otros datos relevantes relacionados con la comprensión profunda del proceso de aprendizaje a efectuar por parte del

profesorado, la combinación deliberada de tareas de carácter acreditativo y pedagógico o la gestión sensible, global y coordinada de las mismas. Entre otras muchas cuestiones, estos focos de atención llevarían a indagar sobre aspectos como la distribución o gestión premeditada del esfuerzo y del trabajo por medio de las demandas evaluativas, los criterios aplicados para la combinación de tareas acreditativas y pedagógicas, el impacto emocional de la evaluación continua, la racionalización de la carga generada en profesorado y estudiantes, la ponderación de las evidencias recopiladas y su transferencia al proceso de calificación o la conciliación de la praxis y la normativa evaluativa (plazos, prerrequisitos, convocatorias, etc.). También sobre las condiciones que permitan avanzar hacia una evaluación del aprendizaje más formadora, favorecedora de la autorregulación del aprendizaje y del desarrollo de la autonomía discente, bases imprescindibles para la consolidación real del principio del *lifelong learning* (Baartman et al., 2013; Bilbao, 2012; Cano & Ion, 2012; Hernández et al., 2006, 2009; Villardón et al., 2013).

- ✓ *La fiabilidad y validez evaluativa desde una perspectiva profunda.* Según se ha constatado a lo largo de la presente investigación, la adecuación y precisión representan áreas de mejora atribuibles a la evaluación. No figuran como núcleos prioritarios de intervención, pero la perspectiva desde la que se abordan en la investigación y las manifiestas discrepancias en las conclusiones comunicadas al respecto hacen que advirtamos en ellas un foco relevante pendiente de profundización. Por lo general, la fiabilidad y la validez se cuestionan por medio de un número reducido de preguntas, centradas en la perspectiva técnica o psicométrica. Sin embargo, la capacidad de la evaluación para valorar su objeto (validez) y la precisión con la que lo hace (fiabilidad) puede estar mediatizada por toda una serie de circunstancias prácticas y pedagógicas, especialmente cuando se pretende hacer de ella un recurso didáctico al servicio del aprendizaje (De la Orden, 2011b; Mateo, 2000). En consecuencia, su exploración debería llevarse a cabo desde una perspectiva más amplia, que vincule ambas propiedades con las funciones y los propósitos de la evaluación y con el resto de decisiones que median en la configuración de los dispositivos evaluativos (cantidad y variedad de evidencias consideradas, cantidad y variedad de instrumentos usados en su recopilación, consistencia interna de la estimación que ofrecen, control de los sesgos y de las condiciones de aplicación, capacidad de los

instrumentos para ofrecer evidencias de las competencias previstas, adecuación a las particularidades del alumnado, etc.).

- *Evaluación y atención a la diversidad.* Los resultados presentados en la presente investigación sugieren un desacuerdo manifiesto al extrapolar el principio de la atención a la diversidad al terreno específico de la evaluación. Valorada su relevancia como indicador de calidad de la praxis evaluativa y la creciente diversificación del perfil discente universitario, consideramos que esta dimensión debería ser objeto de una exploración más profunda. Contemplada también entre los principales ejes de actuación señalados desde el CRESST (2012), su desarrollo debería atender, al menos, a dos niveles de reflexión complementarios:

- ✓ *Concepciones y actitudes que suscita en estudiantes y profesorado la diversificación evaluativa.* La mera presentación de principios o razones para adecuar los dispositivos evaluativos a las necesidades y particularidades discentes no garantiza su transferencia, de modo que es importante detectar las creencias que dan lugar a las resistencias y reticencias que sugieren los resultados evidenciados.
- ✓ *Límites y posibilidades para el ejercicio de la diversificación en el marco de la normativa evaluativa y de las profesiones reguladas.* Al hilo de las dificultades y aparentes contradicciones sugeridas en estudios previos (Ibarra & Rodríguez, 2010; Padilla et al., 2010), debería emprenderse un examen profundo sobre las posibilidades reales que ofrece la normativa evaluativa para la adecuación de los dispositivos evaluativos y sobre la definición de justicia que subyace a la misma. Esta reflexión resulta especialmente relevante en el marco de las profesiones reguladas, sujetas a requisitos normativos en cuanto a la acreditación de las competencias profesionales promovidas.

- *Formación docente para el desarrollo de la competencia evaluativa y transferencia práctica.* Señalada como principal sugerencia de mejora, la formación docente en evaluación representa un importante núcleo de atención. La incipiente línea de investigación que aborda la transferencia de dicha formación al proceso de enseñanza y el aprendizaje discente (Feixas et al., 2013; Monereo, 2013; Prebble, 2004; Southwell & Morgan, 2009) resulta particularmente interesante en este sentido. Su desarrollo debería contribuir a profundizar en aquellas condiciones que mejor favorecen el impacto real en las aulas y el cambio en las concepciones evaluativas del profesorado y las/os futuras/os maestras/os.

- *La competencia evaluativa como resultado de aprendizaje en la formación básica de maestras/os.* Sensible a los procesos de reproducción que tienen lugar en la configuración de las concepciones evaluativas profundas, la presente investigación ha prestado atención a la competencia evaluativa del profesorado vinculado a los Grados de Educación Infantil y Primaria. Sin embargo, la mejora de la praxis evaluativa en el sistema educativo no debería desconsiderar la capacitación evaluativa ofrecida a las/os futuras/os maestras/os a lo largo de su formación básica (Gullickson, 1993/2012; Mertler, 2009; Sanders & Vogel, 1993/2012; Stiggins, 1993/2012, 1998; Stiggins & Conklin, 1988). Los decretos reguladores de las titulaciones que habilitan para el ejercicio de la enseñanza en Educación Infantil y Primaria (Orden ECI/3854/2007 y Orden ECI/3857/2007, respectivamente), incorporan con carácter normativo competencias vinculadas a la práctica evaluativa. Partiendo de dicha base, consideramos pertinente analizar el grado de concordancia de las mismas con la teoría de la evaluación definida (modelo ITC) y el desarrollo que encuentran en los planes de estudio aprobados en las universidades del estado. Este mismo análisis podría replicarse para el Máster de Educación que da acceso a la enseñanza secundaria.
- *Potencialidades de las tecnologías de la información y la comunicación en la evaluación del aprendizaje.* La investigación desarrollada ha rehusado deliberadamente la adopción de un enfoque instrumental en el análisis de las prácticas evaluativas. En consecuencia, no ha prestado atención expresa a las herramientas y recursos tecnológicos y a su aplicabilidad en el desempeño de las funciones evaluativas. Por su utilidad en la gestión, la docencia, la investigación y su transferencia, el manejo de la tecnología en la formación ha pasado a formar parte de las competencias atribuidas al profesorado universitario (Cebrián, 2003; Unigarro & Rondón, 2005; Villa, 2004; Zabalza, 2003). En el contexto específico de la evaluación del aprendizaje, las tecnologías de la información y la comunicación representan un recurso potencialmente efectivo para contrarrestar la sobrecarga, una de las principales dificultades atribuidas a la evaluación. Su aplicación podría ser particularmente beneficiosa para favorecer el seguimiento en el contexto de la evaluación continua (Coll et al., 2007) y contribuir a la mejora de los procesos de retroalimentación (Evans, 2013; Lu & Law, 2012). También para avanzar en el empleo de formatos de evaluación contextualizada, por medio de recursos que faciliten la simulación de tareas reales (Bezanilla et al., 2014; Guenaga et al., 2013).

REFERENCIAS BIBLIOGRÁFICAS

REFERENCIAS BIBLIOGRÁFICAS

A

African Evaluation Association (AfrEA). (2007). *African Evaluation Guidelines-Standards and norms* (version 06/07). Recuperado de <http://www.ader-evaluate.ro/docs/African%20Evaluation%20Association.pdf>

Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA). (2006). *DOCENTIA (Programa de apoyo para la evaluación de la actividad docente del profesorado universitario): Modelo de evaluación*. Recuperado de http://www.aneca.es/var/media/215769/docentia_modelo_070302.pdf

Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA). (2007). *Programa de evaluación del profesorado para la contratación. Guía de ayuda al solicitante*. Recuperado de <http://www.aneca.es/Programas/PEP/Documentos-de-ayuda>

Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA), Àgència per a la Qualitat del Sistema Universitari a Catalunya (AQU), & Axencia para a Calidade do Sistema Universitario de Galicia (ACSUC). (2007a). *Programa AUDIT. Guía para el diseño de sistemas de garantía interna de la calidad de la formación universitaria* (Documento 01). Recuperado de <http://www.aneca.es/Programas/AUDIT/Herramientas-para-el-diseno>

Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA), Àgència per a la Qualitat del Sistema Universitari a Catalunya (AQU), & Axencia para a Calidade do

- Sistema Universitario de Galicia (ACSUC). (2007b). *Programa AUDIT. Directrices, definición y documentación de Sistemas de Garantía Interna de Calidad de la formación universitaria* (Documento 02). Recuperado de <http://www.aneca.es/Programas/AUDIT/Herramientas-para-el-diseno>
- Àgencia per a la Qualitat del Sistema Universitari a Catalunya (AQU). (2003). *Marc general per a l'avaluació dels aprenentatges dels estudiants*. Recuperado de http://www.aqu.cat/doc/doc_42884456_1.pdf
- Ahumada, P. (2002). Estrategias y procedimientos para una evaluación auténtica de los aprendizajes en la enseñanza universitaria. *Cuadernos IRC- Planificación, Gestión y Evaluación Universitaria*, 10, 28-39. Recuperado de http://prontus.uv.cl/pubacademica/pubcarreras/pubmedicinav/site/artic/20060905/asocfile/evalaucion_autentica.pdf
- Ahumada, P. (2005). La evaluación auténtica: Un sistema para la obtención de evidencias y vivencias de aprendizaje. *Perspectiva Educacional*, 45, 11-24. Recuperado de <http://www.redalyc.org/pdf/3333/333329100002.pdf>
- Albert, M. J. (2009). *La investigación educativa. Claves teóricas*. Madrid: McGraw-Hill.
- Alegre, O. M., & Villar, L. M. (2006). *Competencias para la formación de los docentes universitarios*. Málaga: Ediciones Aljibe.
- Alfaro, I. J. (2006). Seminarios y talleres. En M. De Miguel (Coord.), *Metodologías de enseñanza y aprendizaje para el desarrollo de competencias. Orientaciones para el profesorado universitario ante el Espacio Europeo de Educación Superior* (pp. 53-81). Madrid: Alianza Editorial.
- Alkin, M. C. (1969). Evaluation theory development. *Evaluation Comment*, 2, 2-7.

- Alsina, J. (2011). Las competencias transversales: Cómo evaluar su aprendizaje. En J. Alsina (Coord.), *Evaluación por competencias en la universidad* (Cuadernos de docencia universitaria, 18) (pp. 18-25). Recuperado de <http://www.octaedro.com/ice/cuadernos.htm>
- Álvarez, I. (2008a). La coevaluación como alternativa para mejorar la calidad del aprendizaje de los estudiantes universitarios: Valoración de una experiencia. *Revista Interuniversitaria de Formación del Profesorado*, 63(22,3), 127-140.
- Álvarez, I. (2008b). Evaluación del aprendizaje en la universidad: Una mirada retrospectiva y prospectiva desde la divulgación científica. *Revista Electrónica de Investigación Psicoeducativa*, 6(14), 235-272. Recuperado http://www.investigacion-psicopedagogica.com/revista/articulos/14/espanol/Art_14_228.pdf
- Álvarez, I. (2009). Evaluar para contribuir a la autorregulación del aprendizaje. *Electric Journal of Research in Educational Psychology*, 7(3), 1007-1030. Recuperado de http://www.investigacion-psicopedagogica.org/revista/articulos/19/english/Art_19_368.pdf
- Álvarez, I., & Olivera-Smith, M. (2014). Autorregulados. Sistemas basados en guías y rúbricas que promueven la autorreflexión. En C. Monereo (Coord.), *Enseñando a enseñar en la universidad. Formación del profesorado basada en incidentes críticos* (pp. 188-216). Barcelona: Octaedro.
- Álvarez, J. M. (2000). *Evaluar para conocer, examinar para excluir*. Madrid: Morata.
- Álvarez, J. M. (2003). *La evaluación a examen: Ensayos críticos*. Buenos Aires: Miño y Dávila.
- Álvarez, R. (1995). *Estadística multivariante y no paramétrica con SPSS: Aplicación a las ciencias de la salud*. Madrid: Díaz de Santos.

- Álvarez, V., Asensio, I., Clares, J., del Frago, R., García, B., García, N., ... Salmerón, P. (2009). Perfiles docentes para el Espacio Europeo de Educación Superior (EEES) en el ámbito universitario español. *Revista Electrónica de Investigación y Evaluación Educativa*, 15(1). Recuperado de http://www.uv.es/RELIEVE/v15n1/RELIEVEv15n1_1.htm
- Álvarez, V., García, E., & Gil, J. (1999). La calidad de la enseñanza universitaria desde la perspectiva de los profesores mejor valorados por los alumnos. *Revista de Educación*, 319, 273-290.
- Álvarez, V., García, E., Gil, J., Romero, S., & Correa, J. (2002). Enseñanza en la universidad: Percepciones de profesores y alumnos. *Revista de Educación*, 328, 303-323.
- Álvarez, I., & Gómez, I. (2009). PISA: Un proyecto internacional de evaluación auténtica. Luces y sombras. En C. Monereo (Coord.), *Pisa como excusa. Repensar la evaluación para cambiar la enseñanza* (pp. 91-110). Barcelona: Graó.
- Álvarez, B., González, C., & García, N. (2007). La motivación y los métodos de evaluación como variables fundamentales para estimular el aprendizaje autónomo. *Revista de Docencia Universitaria*, 5(2). Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/53/pdf>
- Álvarez, V., Romero, S., Gil, J., Rodríguez, J., Clares, J., Asensio, I., ... Salmeron, P. (2011). Necesidades de formación del profesorado universitario para la adaptación de su docencia al Espacio Europeo de Educación Superior (EEES). *Revista Electrónica de Investigación y Evaluación Educativa*, 17(1). Recuperado de http://www.uv.es/RELIEVE/v17n1/RELIEVEv17n1_1.htm
- American Association of School Administrators, National Association of Elementary School Principals, National Association of Secondary School Principals, & National Council

on Measurement in Education. (1997). *Competency standards in student assessment for educational administrators*. Recuperado de <http://buros.org/competency-standards-student-assessment-educational-administrators>

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. Washington: Autor.

American Evaluation Association (AEA). (2004). *American Evaluation Association guiding principles for evaluators*. Recuperado de <http://www.eval.org/p/cm/ld/fid=51>

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30-32. doi: 10.1111/j.1745-3992.1990.tb00391.x

American Psychological Association (APA). (2010). *Manual de Publicaciones de la American Psychological Association* (6ª ed.). México: Manual Moderno.

American Public University System (APUS). (2013). *Glossary of assessment terms*. Recuperado de <http://www.apus.edu/community-scholars/learning-outcomes-assessment/university-assessment/glossary.htm>

Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-18.

Andrade, H. (2005). Teaching with rubrics: The good, the bad and the ugly. *College Teaching*, 53(1), 27-30. doi: 10.3200/CTCH.53.1.27-31

- Andresen, L. W. (2000). A useable, trans-disciplinary conception of scholarship. *Higher Education Research and Development*, 19(2), 137-153. doi: 10.1080/072943600445619
- Andreu-Andrés, M. A., & Labrador-Piquer, M. J. (2011). Formación del profesorado en metodologías y evaluación. Análisis cualitativo. *Revista de Investigación en Educación*, 9(2), 236-245. Recuperado de <http://webs.uvigo.es/reined/ojs/index.php/reined/article/view/303/170>
- Aparicio, J. A., & Pozo, J. I. (2006). De fotógrafos a directores de orquesta: Las metáforas desde las que los profesores conciben el aprendizaje. En J. I. Pozo et al. (Eds.), *Nuevas formas de pensar la enseñanza y el aprendizaje. Concepciones de profesores y alumnos* (pp. 265-288). Barcelona: Graó.
- Apodaca, P., & Grao, J. (1997). Herramientas de gestión para el cambio y la mejora institucional en la enseñanza superior. En P. Apodaca & J. Grao (Eds.), *Calidad en la universidad: Orientación y evaluación* (pp. 175-193). Barcelona: Laertes.
- Arénaga, S. (2013). *De la detección de las necesidades de formación pedagógica a la elaboración de un plan de formación en la universidad* (Cuadernos de docencia universitaria, 25). Recuperado de <http://www.ub.edu/ice/sites/default/files//docs/qdu/25cuaderno.pdf>
- Arizona's State Board for Education (ASBE). (1996). *Arizona's professional teacher standards*. Recuperado de http://swmcdn.com/site_0157/UnionHurleyRanch_AZProfTeachStandards_080514.pdf
- Arribas, J. M. (2012). El rendimiento académico en función del sistema de evaluación empleado. *Revista Electrónica de Investigación y Evaluación Educativa*, 18(1). Recuperado de http://www.uv.es/RELIEVE/v18n1/RELIEVEv18n1_3.pdf

- Arribas, J. M., Carabias, D., & Monreal, I. (2010). La docencia universitaria en la formación inicial del profesorado. El caso de la escuela de magisterio de Segovia. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 13(3), 27-25.
Recuperado de http://aufop.com/aufop/uploaded_files/articulos/1285861727.pdf
- Arter, J. (abril, 2009). *Classroom Assessment for Student Learning (CASL). Perspective on the JCSEE Student Evaluation Standards*. Paper presentado en el encuentro anual de la American Educational Research Association (AERA). Education research in the circle of knowledge. San Diego. Recuperado de <http://ati.pearson.com/downloads/JCSEE%20Student%20Eval%20Practices%20Symposium%20Paper%202009.pdf>
- Assessment Reform Group (ARG). (1999). *Assessment for learning: Beyond the black box*. Recuperado de <http://www.aaia.org.uk/content/uploads/2010/06/Assessment-for-Learning-Beyond-the-Black-Box.pdf>
- Assessment Reform Group (ARG). (2002). *Assessment for learning: 10 principles. Research-based principles to guide classroom practice*. Recuperado de http://assessmentreformgroup.files.wordpress.com/2012/01/10principles_english.pdf
- Assessment Resource at Hong Kong University (AR@HKU). (2009a). *Assessment Methods*. Recuperado de http://ar.cetl.hku.hk/assessment_method.htm
- Assessment Resource at Hong Kong University (AR@HKU). (2009b). *Glossary*. Recuperado de <http://ar.cetl.hku.hk/terminologies.htm>
- Assessment Resource at Hong Kong University (AR@HKU). (2009c). *OBASL (Outcomes Based Approach to Student Learning) & relationship with learning outcomes, learning activities and assessment*. Recuperado de <http://ar.cetl.hku.hk/obasl.htm>

Assessment Resource at Hong Kong University (AR@HKU). (2009d). *22 tips for designing an effective assessment*. Recuperado de <http://ar.cetl.hku.hk/tips.htm>

Astin, A., Banta, T., Cross, P., El-Khawas, E., Ewell, P., Hutchings, P., ... Wright, B. (1992). *Nine principles of good practice for assessing student learning*. Recuperado del sitio de Internet de la Association of Higher Education (AAHE):
http://www.academicprograms.calpoly.edu/pdfs/assess/nine_principles_good_practice.pdf

Atkin, J. M, Black, P., & Coffey, J. (2001). *Classroom assessment and the National Science Standards*. Washington, DC: National Academy Press.

Atkinson, D., & Lim, S. L. (2013). Improving assessment processes in Higher Education: Student and teacher perceptions of the effectiveness of a rubric embedded in a LMS. *Australasian Journal of Educational Technology*, 29(5), 651-666. Recuperado de
http://espace.library.curtin.edu.au/R?func=dbin-jump-full&local_base=gen01-era02&object_id=193744

Australian Institute for Teaching and School Leadership (AITSL). (2012). *Australian professional standards for teachers*. Recuperado de
<http://www.aitsl.edu.au/Pages/Standards/Pdf.aspx?&s=1&s=2&s=3&s=4&s=5&s=6&s=7>

B

Baartman, L. K., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C. P. (2006). The wheel of competency assessment. Presenting quality criteria for competency assessment programs. *Studies in Educational Evaluation*, 32, 153-177. doi: 10.1016/j.stueduc.2006.04.006

Baartman, L. K., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C. P. (2007).

Evaluation assessment quality in competence-based education: A qualitative comparison of two frameworks. *Educational Research Review*, 2, 114-129. doi: 10.1016/j.edurev.2007.06.001

Baartman, L. K., Gulikers, J., & Dijkstra, A. (2013). Factors influencing assessment quality in higher vocational education. *Assessment & Evaluation in Higher Education*, 38(8), 978-997. doi: 10.1080/02602938.2013.771133

Badía, A., & Álvarez, I. (2009). Enseñar y aprender en la educación superior a través de la evaluación auténtica en entornos virtuales de aprendizaje. En M. Castelló (Coord.), *La evaluación auténtica en enseñanza secundaria y universitaria: Investigación e innovación* (pp. 171-192). Barcelona: Edebé.

Baeza, J., Pérez, M., & Reyes, L. (2006). *Estándares de desempeño docente, sistematización de contenidos para su operacionalización*. Recuperado de http://www.oei.es/docentes/articulos/estandares_desempeno_docente_reyes.pdf

Bain, K. (2007). *Lo que hacen los mejores profesores universitarios* (2ª ed.). Universidad de Valencia.

Baker, E., Linn, R. L., & Herman, J. L. (1996). *Evaluation comment: A continuing mission to improve the educational assessment*. Recuperado del sitio de Internet del National Center for Research on Evaluation, Standards & Student Testing (CRESST): http://www.cse.ucla.edu/products/evaluation/cresst_ec1996_3.pdf

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education*, 27(5), 427-441. doi: 10.1080/0260293022000009302

- Banta, T. W. (2007). Can assessment for accountability complement assessment for improvement? *Peereview*, 9(2), 9-12.
- Banta, T. W., Lund, J. P., Black, K. E., & Oblander, F. W. (1996). *Assessment in practice: Putting principles to work on college campuses*. San Francisco: Jossey-Bass.
- Barberá, E. (1999). *Evaluación de la enseñanza, evaluación del aprendizaje*. Barcelona: Edebé.
- Barberá, E., & Martín, E. (2009). *Portfolio electrónico: Aprender a evaluar el aprendizaje*. Barcelona: Editorial UOC.
- Barnett, R. (2005). *Reshaping the university: New relationships between research, scholarship and teaching*. Maidenhead: Open University Press.
- Barnett, R. (2008). Introducción. En R. Barnett (Ed.), *Para una transformación de la universidad. Nuevas relaciones entre investigación, saber y docencia* (p. 13-19). Barcelona: Octaedro.
- Barr, R. B., & Tagg, J. (1995). From teaching to learning. A new paradigm for undergraduate education. *Change*, 27(6), 13-25. Recuperado de http://cet.usc.edu/resources/teaching_learning/docs/teaching_to_learning.pdf
- Barth, B. M. (2004). La evaluación a la luz de un enfoque socio-constructivista del aprendizaje. En J. C. Torre & E. Gil (Eds.), *Hacia una enseñanza universitaria centrada en el aprendizaje. Libro homenaje a Pedro Morales* (pp. 257-269). Madrid: Universidad Pontificia de Comillas.
- Bassani, A. (2009). *Estrategias didácticas semipresenciales mediadas por las tecnologías de la información y la comunicación* (Tesis doctoral, Universidad Rovira i Virgili- Universidad Internacional de Salta). Recuperado de <http://www.tdx.cat/bitstream/handle/10803/8946/BASSANI.pdf?sequence=1>

- Blanch, S., & Corcelles, M. (2014). Reciprocidades. Sistemas de formación basados en la cooperación entre docentes. En C. Monereo (Coord.), *Enseñando a enseñar en la universidad. Formación del profesorado basada en incidentes críticos* (pp. 156-187). Barcelona: Octaedro.
- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation*, 18(6). Recuperado de <http://pareonline.net/pdf/v18n6.pdf>
- Bejarano, M. (2008). Modelos tradicionales y nuevos modelos para una enseñanza universitaria enmarcada en el Espacio Europeo de Educación Superior (EEES). *Multiárea*, 3, 27-38.
- Belmonte, M. (2002). *Enseñar a investigar. Orientaciones prácticas*. Bilbao: Ediciones Mensajero.
- Beltrán, J., & Pérez, L. F. (2005). El profesor universitario desde la perspectiva de los alumnos. *Revista de Psicología y Educación*, 1, 79-114. Recuperado de <http://revistadepsicologiayeducacion.es/index.php/descargasi/finish/3/8.html>
- Benito, A., & Cruz, A. (2005). Introducción. En A. Benito & A. Cruz (Coords.), *Nuevas claves para la docencia universitaria en el EEES* (pp. 11-20). Madrid: Narcea.
- Berk, R. A. (2005). Survey of 12 Strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17(1), 48-62. Recuperado de <http://www.isetl.org/ijtlhe/pdf/IJTLHE8.pdf>
- Berlanga, V., & Rubio, M. J. (2012). Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS. *Revista d'Innovació i Recerca en Educació*, 5(2), 101-113. doi: 10.1344/reire2012.5.2528

- Bevitt, S. (2015). Assessment innovation and student experience: A new assessment challenge and call for a multi-perspective approach to assessment research. *Assessment & Evaluation in Higher Education*, 40(1), 103-119. doi: 10.1080/02602938.2014.890170
- Bezanilla, M. J., Arranz, S., Rayón, A., Rubio, I., Mentxaka, I., Guenaga, M., & Aguilar, E. (2014). A proposal for generic competence assessment in a serious game. *Journal of New Approaches in Educational Research*, 3(1), 42-51. doi: 10.7821/naer.3.1.42-51
- Biggs, J. (2005). Calidad del aprendizaje universitario. Madrid: Narcea.
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Maidenhead: Open University Press-McGraw-Hill.
- Bilbao, A. (2010). La evaluación como medio y como fin. Una experiencia de uso estratégico de la evaluación en Didáctica General. En J. Marrero & J. Carbonell (Coords.), *Actas del I Congreso internacional Reinventar la profesión docente* (Mesa 10) (pp. 70-88). Universidad de Málaga. Recuperado de <http://www.doe.uma.es/repository/fileDownloader?rfname=2b8ec5ec-bd31-4748-b78d-720373122bf0.pdf>
- Bilbao, A. (2011a). Teoría y práctica del valor estratégico de la evaluación del aprendizaje en la formación básica de maestros: Una experiencia de metaevaluación a cargo de estudiantes. En Universidad de Deusto, Universidad Pontificia de Comillas, & Universidad Ramón Llull, *Jornadas Interuniversitarias de Innovación Docente* (pp. 167-174). Barcelona: Blanquerna Tecnologia i Serveis -Universitat Ramon Llul. Recuperado de <http://www.aristoscampusmundus.net/wp-content/uploads/2013/05/JornadasInteruniversitarias.pdf>
- Bilbao, A. (2011b). El uso estratégico de la evaluación en el Espacio Europeo de Educación Superior (EEES). En M. Cerezo (Dir.), *III Jornada Nacional sobre Estudios*

- Universitarios. El presente de los nuevos títulos* (pp. 109-117) [CD]. Castellón: Universidad Jaume I.
- Bilbao, A. (2012). La evaluación del aprendizaje como recurso estratégico en la promoción de la autonomía del alumnado universitario: Paradoja de la hiperregulación para la autorregulación futura. En A. Villa (Ed.), *Avances en la innovación universitaria. Tejiendo el compromiso de las universidades* (pp. 525-537). Bilbao: Ediciones Mensajero.
- Bilbao, A. (2014). Uso estratégico de la evaluación del aprendizaje en la formación básica de maestros. En M. M. Martín, E. Reche, & I. González (Coords.), *La evaluación de competencias. Innovación en las aulas universitarias* (pp. 133-146). Córdoba: Centro de Magisterio Sagrado Corazón.
- Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. En M. Birenbaum & F. Dochy (Eds.), *Alternative in assessment of achievement, learning processes and prior knowledge* (pp. 3-29). Boston: Kluwer.
- Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., ... Nickmans, G. (2006). A learning integrated assessment system. *Educational Research Review*, 1(1), 61-67. doi: 10.1016/j.edurev.2006.01.001
- Biscarri, J., Filella, G., & Jové, G. (2006). Factores relacionados con la percepción de calidad docente del profesorado universitario. *Revista Interuniversitaria de Formación del Profesorado*, 20(3), 287-309.
- Bisquerra, R. (1987). *Introducción a la estadística aplicada a la investigación educativa*. Barcelona: Promociones Publicaciones Universitarias.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, 4(1), 7-71. doi: 10.1080/0969595980050102

- Black, P., & Wiliam, D. (1998b). Inside the black box. *Phi Delta Kappan*, 80(2), 139-148.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5-31. doi: 10.1007/s11092-008-9068-5
- Blanco, A. (2008a). Formación universitaria basada en competencias. En L. Prieto (Coord.), *La enseñanza universitaria centrada en el aprendizaje. Estrategias útiles para el profesorado* (pp. 31-59). Barcelona: Octaedro-ICE UB.
- Blanco, A. (2008b). Las rúbricas: Un instrumento útil para la evaluación de competencias. En L. Prieto (Coord.), *La enseñanza universitaria centrada en el aprendizaje. Estrategias útiles para el profesorado* (pp. 171-188). Barcelona: Octaedro-ICE UB.
- Blanco, A., & Alba, E. (2009). La participación de los estudiantes y el trabajo en equipo. En A. Blanco (Coord.), *Desarrollo y evaluación de competencias en educación superior* (pp. 49-63). Madrid: Narcea.
- Bloxham, S. (2013). Building `standards´framework: The role of guidance and feedback in supporting the achievement of learners. En S. Merry, M. Price, D. Carless, & M. Taras (Eds.), *Reconceptualising feedback in higher education: Developing dialogue with students* (pp. 64-74). London: Routledge.
- Bolarín, M., Moreno, M., & Porto, M. (2013). Coordinación docente e interdisciplinariedad: Análisis de su contribución a la adquisición de competencias docentes y discentes. *Revista de Docencia Universitaria*, 11(2), 443-462. Recuperada de <http://red-u.net/redu/index.php/REDU/article/view/455>
- Bonsón, M. (2009). Desarrollo de competencias en educación superior. En A. Blanco (Coord.), *Desarrollo y evaluación de competencias en educación superior* (pp. 17-34). Madrid: Narcea.

- Bonsón, M., & Benito, A. (2005). Evaluación y aprendizaje. En A. Benito & A. Cruz (Coords.), *Nuevas claves para la docencia universitaria en el EEES* (pp. 87-100). Madrid: Narcea.
- Bordas, M. I., & Cabrera, F. (2001): Estrategias de evaluación de los aprendizajes centrados en el proceso. *Revista española de pedagogía*, 218, 25-48.
- Boud, D. (1995). Introduction. En D. Boud (Coord.), *Enhancing learning through self-assessment* (1-10). London: Routledge Falmer.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167. doi: 10.1080/01580379950177314
- Boud, D., & Asociados. (2010). *Assessment 2020. Seven propositions for assessment reform in higher education*. Recuperado de:
http://www.uts.edu.au/sites/default/files/Assessment-2020_propositions_final.pdf
- Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment & Evaluation in Higher Education*. 24(4), 413-426. doi: 10.1080/0260293990240405
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399-413. doi: 10.1080/02602930600679050
- Boud, D., & Falchikov, N. (2007). Introduction: Assessment for long term. En D. Boud & N. Falchikov (Ed.), *Rethinking assessment in higher education. Learning for the longer term* (pp. 3-14). London: Routledge.
- Boud, D., & Molloy, E. (2013). What is the problem with the feedback? En D. Boud & E. Molloy (Ed.), *Effective feedback in higher and professional education: Understanding it and doing it well* (pp. 1-10). Abingdon: Routledge.
- Boyer, E. (1990). *Scholarship reconsidered*. Washington, DC: The Carnegie Foundation.

- Bozu, Z., & Canto, P. J. (2009). El profesorado universitario en la sociedad del conocimiento: Competencias profesionales docentes. *Revista de Formación e Innovación Educativa Universitaria*, 2(2), 87-97. Recuperado de http://webs.uvigo.es/refiedu/Refiedu/Vol2_2/REFIEDU_2_2_4.pdf
- Bretones, A. (2008). Participación del alumnado de educación superior en su evaluación. *Revista de Educación*, 347, 181-202. Recuperado de http://www.revistaeducacion.mec.es/re347/re347_09.pdf
- Brew, A. (2003). La autoevaluación y la evaluación por los compañeros. En S. Brown & A. Glasner (Eds.), *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 179-189). Madrid: Narcea.
- Brew, A. (2007). Integrating research and teaching. Understanding excellence. En A. Skelton (Ed.), *International perspectives on teaching excellence in higher education. Improving knowledge and practice* (pp. 74-88). Canada: Routledge.
- Brew, C., Riley, P., & Walta, C. (2009). Education students and their teachers: Comparing views on participative assessment practices. *Assessment & Evaluation in Higher Education*, 34(6), 641-657. doi: 10.1080/02602930802468567
- Briggs, C. L., Stark, J. S., & Rowland-Poplowsky, J. (2003). How do we know a 'continuous planning' academic program when we see one? *The Journal of Higher Education*, 74(4), 361-385. doi: 10.1353/jhe.2003.0028
- Brockbank, A., & McGill, I. (2002). *Aprendizaje reflexivo en la educación superior*. Madrid: Morata.
- Brodie, D. A. (1998). Do students report that easy professors are excellent teachers? *Canadian Journal of Higher Education*, 23(1), 1-20. Recuperado de la base de datos ERIC. (EJ565303)

- Brown, S. (2003a). Aplicaciones prácticas de una evaluación práctica. En S. Brown & A. Glasner (Eds.), *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 117-127). Madrid: Narcea.
- Brown, S. (2003b). Estrategias institucionales en evaluación. En S. Brown & A. Glasner (Eds.), *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 23-33). Madrid: Narcea.
- Brown, S., Bull, J., & Pendlebury, M. (1997). *Assessing students learning in higher education*. London. Routledge.
- Brown, S., & Knight, P. (1994). *Assessing learners in higher education*. London: Kogan Page.
- Brown, S., Race, P., & Smith, B. (1996). *500 Tips on Assessment*. London: Kogan Page.
- Bryson, C., McDowell, L., McGugan, S., & Sanders, G. (2008). *A relationship heading for divorce? Assessment, trust and student engagement*. Recuperado de http://www.brookes.ac.uk/aske/documents/bryson_mcdowell.pdf
- Buckley, A. (2012). *National Student Survey. Analysis of national results for 2011*. Recuperado del sitio de Internet de la Higher Education Academy (HEA): https://www.heacademy.ac.uk/sites/default/files/NSS_report_2011_FINAL.pdf
- Buela-Casal, G., Bermúdez, M. P., Sierra, J. C., Quevedo-Blasco, R., & Guillén-Riquelme, A. (2012). Ranking 2012 de investigación de las universidades públicas españolas. *Psicothema*, 26(2), 149-158. doi: 10.7334/psicothema2013.312
- Buendía, L., Carmona, M., González, D., & López, R. (1999). Concepciones de los profesores de educación secundaria sobre evaluación. *Educación XX1*, 2, 125-153. Recuperado de <http://revistas.uned.es/index.php/educacionXX1/article/view/377/328>

Buendía, L., & López, R. (1999). Concepciones de los profesores universitarios sobre la evaluación de los alumnos. *Organización y Gestión Educativa*, 6, 19-30.

Burgess, R. (Dir.). (2004). *Measuring and recording students achievement* (Report of the Scoping Group chaired by Professor Robert Burgess). Recuperado del sitio de Internet de Universities UK.:
<http://www.universitiesuk.ac.uk/highereducation/Documents/2004/MeasuringAchievement.pdf>

Burgess, R. (Dir.). (2007). *Beyond the honors degree classification* (The Burgess Group final report). Recuperado del sitio de Internet de Universities UK.:
<http://www.universitiesuk.ac.uk/highereducation/Documents/2007/BurgessBeyondHonours.pdf>

Burgess, R. (Dir.). (2012). *Bringing it all together: Introducing the HEAR* (The final report of the Burgess Implementation Steering Group). Recuperado del sitio de Internet de Universities UK: <http://www.hear.ac.uk/assets/documents/hear/institution-resources/HEAR-Bringing-it-all-together.pdf>

Buscà, F., Capllonch, M., Cladellas, L., Calvo, J., Martín, M., & Padrós, M. (2010). Sistemas de evaluación formativa y participativa en docencia universitaria: Un estudio de casos aplicado a la formación del profesorado. En T. Pagès, A. Cornet, & J. Pardo (Coords.), *Buenas prácticas docentes en la universidad. Modelos y experiencias en la Universidad de Barcelona* (pp. 44-53). Barcelona: Octaedro-ICE UB.

Buscà, F., Pintor, P., Martínez, L., & Peire, T. (2010). Sistemas y procedimientos de evaluación formativa en docencia universitaria: Resultados de 34 casos aplicados durante el curso académico 2008-2009. *Estudios Sobre Educación*, 18, 255-276. Recuperado de http://dadun.unav.edu/bitstream/10171/9829/2/ESE_18_11.pdf

C

- Caballero, K. (2013). La formación del profesorado universitario y su influencia en el desarrollo de la actividad profesional. *Revista de Docencia Universitaria*, 11(2), 391-412. Recuperado de <http://red-u.net/redu/files/journals/1/articles/446/public/446-2381-1-PB.pdf>
- Cabra, F. (2005). *Calidad y estándares de la evaluación del estudiante* (Tesis doctoral). Universidad de Deusto, Bilbao.
- Cabra, F. (2008). La calidad de la evaluación de estudiantes: Un análisis desde estándares profesionales. *Revista Internacional de Investigación en Educación*, 1, 95-112. Recuperado de <http://www.redalyc.org/pdf/2810/281021687007.pdf>
- Cabrera, A., & La Nasa, S. M. (2005). Classroom teaching practices: Ten lessons learned. En W. De Vries (Coord.), *Calidad eficiencia y evaluación de la educación superior* (pp. 129-158). La Coruña: Netbiblo.
- Calatayud, A. (2008). Si cambia la metodología docente ha de cambiar la evaluación en la universidad. Hacia nuevas formas de evaluación. En *Actas del V Congreso Internacional Docencia Universitaria e Innovación: El cambio en la cultura docente universitaria* [CD]. Universidad de Lleida.
- Calatayud, M. A. (2004). Reflexión de los alumnos de primaria sobre preconcepciones evaluativas. En A. Parcerisa (Dir.), *Evaluación como ayuda al aprendizaje. Claves para la innovación educativa* (pp. 91-102). Barcelona: Graó.
- Calbó, M. (Coord.). (2009). *Guía para la evaluación de competencias en el prácticum de los estudios de maestra/o*. Recuperado del sitio de Internet de la Agència per a la

Qualitat del Sistema Universitari de Catalunya (AQU):

http://www.aqu.cat/doc/doc_84811405_1.pdf

Camargo, I. M., & Pardo, C. (2009). Competencias docentes de profesores de pregrado:

Diseño y validación de un instrumento de evaluación. *Universitas Psychologica*, 7(2), 441-455. Recuperado de

<http://revistas.javeriana.edu.co/index.php/revPsycho/article/view/441>

Camilloni, A. (1998). La calidad de los programas de evaluación y de los instrumentos que

los integran. En A. Camilloni, et al. (Eds.), *La evaluación de los aprendizajes en el debate didáctico contemporáneo* (pp. 67-92). Buenos Aires: Paidós Educador.

Campbell, D. (2012). Teachers' assessment of students: Roles, responsibilities and purpose

believing is seeing. En S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 97-128). Recuperado de

<http://digitalcommons.unl.edu/burosteachertraining/> (Trabajo original publicado en 1993).

Campos, J., Ligia, O., & Solano, A. (2010). Competencias del docente de educación

superior en línea. *Revista Electrónica Actualidades Investigativas en Educación*, 10(3). Recuperado de <http://www.redalyc.org/pdf/447/44717980010.pdf>

Cano, E. (2007). *Cómo mejorar las competencias de los docentes. Guía para la*

autoevaluación y el desarrollo de las competencias del profesorado (2ª ed.).

Barcelona: Graó.

Cano, E. (Coord.). (2011). *Buenas prácticas en la evaluación de competencias. Cinco casos*

en educación superior. Barcelona: Laertes.

Cano, E. (2014). Análisis de las investigaciones sobre feedback: Aportes para su mejora en

el marco del EEES. *Bordón*, 66(4), 9-24. doi: 10.13042/Bordon.2014.66402

Cano, E., & Ion, G. (2012). Prácticas evaluadoras en las universidades catalanas: Hacia un modelo centrado en competencias. *Estudios Sobre Educación*, 22, 155-177.

Recuperado de <http://dadun.unav.edu/handle/10171/22627>

Cano, E., Lleixà, T., Ion, G., & Tierno, J. M. (2010). Criterios de una buena práctica en evaluación. En T. Pagès, A. Cornet, & J. Pardo (Coords.), *Buenas prácticas docentes en la universidad. Modelos y experiencias en la Universidad de Barcelona* (pp. 86-97). Barcelona: Octaedro-ICE UB.

Capllonch, M., & Buscà, F. (2012). La evaluación formativa como innovación. Experiencias en una facultad de formación del profesorado. *Psychology, Society, & Education*, 4(1), 45-58. Recuperado de <http://www.psye.org/articulos/vista%20Capllonch%20Bujosa.pdf>

Capllonch, M., Ureña, N., Ruiz, E., González, N., Pérez, A., López, V. M., ... Tabernero, B. (2009). Guía de problemas y soluciones para el desarrollo de sistemas de evaluación formativa en la universidad. En V. M. López (Coord.), *La evaluación formativa y compartida en docencia universitaria: Propuestas, técnicas, instrumentos y experiencias* (pp. 217-252). Madrid: Narcea.

Cardona, C. (2002). *Introducción a los métodos de investigación en educación*. Madrid: EOS.

Care, N. (1978). Participation and policy. *Ethics*, 88, 316-337.

Carless, D. (2002). The mini-viva as a tool to enhance assessment for learning. *Assessment & Evaluation in Higher Education*, 27(4), 353-363. doi: 10.1080/0260293022000001364

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219-233. doi: 10.1080/03075070600572132

- Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57-66. doi: 10.1080/14703290601081332
- Carless, D. (2013a). Sustainable feedback and the development of student self-evaluative capacities. En S. Merry, M. Price, D. Carless, & M. Taras (Eds.), *Reconceptualising feedback in higher education. Developing dialogue with students* (pp. 117-126). London: Routledge.
- Carless, D. (2013b). Trust and its role in facilitating dialogic feedback. En D. Boud & L. Molloy (Eds.), *Effective feedback in higher and professional education: Understanding it and doing it well* (pp. 90-124). London: Routledge.
- Carless, D., Joughin, G., & Mok, M. C. (2006). Learning-oriented assessment: Principles and practice. *Assessment & Evaluation in Higher Education*, 31(4), 395-398. doi: 10.1080/02602930600679043
- Carless, D., Joughin, G., Liu, N. F., & Associates. (2006). *How assessment supports learning: Learning-oriented assessment in action*. Hong Kong University Press.
- Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education*, 36, 395-407. doi: 10.1080/03075071003642449
- Carmena, G., Asensio, M., De la Cuesta, J. C., Doncel, D., & Rodríguez, M. (2005). *Criterios y procedimientos de evaluación de la investigación educativa*. Madrid: Ministerio de Educación y Ciencia-CIDE.
- Carretero, M. (1994). *Constructivismo y educación*. Zaragoza: Edelvives.
- Casanova, M. A. (1992). *La evaluación garantiza de calidad para el centro educativo*. Zaragoza. Edelvives.

- Casanova, M. A. (1998). *La evaluación educativa. Escuela básica*. Madrid: SEP-La Muralla.
- Castejón, F. J., Capllonch, M., González, N., & López, V. M. (2009). Técnicas e instrumentos de evaluación. En V. M. López (Coord.), *Evaluación formativa y compartida en educación superior. Propuestas, técnicas, instrumentos y experiencias* (pp. 65-91). Madrid: Narcea.
- Castejón, F. J., Santos, M. L., & Palacios, A. (en prensa). Cuestionario sobre metodología y evaluación en formación inicial en educación física. *Revista Internacional de Medicina y Ciencias de la Actividad y el Deporte*. Recuperado de <http://cdeporte.rediris.es/revista/inpress/artescala566.pdf>
- Castejón, F. J., López, V. M., Julián, J. A., & Zaragoza, J. (2011). Evaluación formativa y rendimiento académico en la formación inicial del profesorado de educación física. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte*, 11(42), 328-346. Recuperado de <http://cdeporte.rediris.es/revista/revista42/artevaluacion163.pdf>
- Castelló, M., Monereo, C., & Gómez, I. (2009). Las competencias de los alumnos y su evaluación. En C. Monereo (Coord.), *Pisa como excusa. Repensar la evaluación para cambiar la enseñanza* (pp. 33-53). Barcelona: Graó.
- Castillo, S. (2004). Use y proporcione retroacción. En L. M. Villar (Coord.), *Programa para la mejora de la docencia universitaria* (pp. 557-573). Madrid: Pearson-Prentice Hall.
- Castillo, S., & Cabrerizo, J. (2003). *Evaluación educativa y promoción escolar*. Madrid: Pearson Educación.
- Caurcel, M. J., & Morales, J. (2008). La enseñanza y el aprendizaje en la universidad. En A. Rodríguez, M. J. Caurcel, & A. M. Ramos (Coords.), *Didáctica en el Espacio Europeo de Educación Superior* (pp. 46-70). Madrid: Editorial EOS.

Cebrián, M. (2008). Los procesos de innovación didáctica en el marco del Espacio Europeo de Educación Superior (EEES). En I. Rodríguez (Ed.), *El nuevo perfil del profesor universitario en el EEES: Claves para la renovación metodológica* (pp. 19-35).

Valladolid: Universidad Europea Miguel de Cervantes.

Center for Assessment & Research Studies (CARS). (2003). *Dictionary of student outcome assessment*. Recuperado de <http://people.jmu.edu/yangsx/>

Centre for Learning Excellence, University of Bedfordshire (CLE-UB). (2011). *A guide to assessment for learning* (Guide nº3). Recuperado de http://www.beds.ac.uk/data/assets/pdf_file/0003/274152/Guide-3-A-Guide-to-Assessment-for-Learning-October-2011.pdf

Centre for the Study of Higher Education (CSHE). (2002). *Core principles of effective assessment. Assessing learning in Australian universities*. Recuperado de <http://www.cshe.unimelb.edu.au/assessinglearning/05/index.html>

Centre for the Study of Higher Education (CSHE). (2010). *Assessing group work*. Recuperado de <http://www.cshe.unimelb.edu.au/assessinglearning/docs/Group.pdf>

Chan, C. (2010). *Assessment: Assessing group work*. Recuperado de <http://ar.cetl.hku.hk/group.htm>

Chappuis, J. (2005). Helping students understand assessment. *Educational Leadership*, 63(3), 39-43.

Chappuis, J. 2009. *Seven strategies of assessment for learning. A study guide from Pearson Assessment Training Institute*. Recuperado de http://ati.pearson.com/downloads/SevenStrategies_StudyGuide_web.pdf

Chappuis, J. (2012). How am I doing? *Educational Leadership*, 70(1), 36-40.

- Chappuis, J. (2014). Thoughtful assessment with the learner in mind. *Educational Leadership*, 71(6), 20-26.
- Chappuis, S., & Chappuis, J. (2007). The best value in formative assessment. *Educational Leadership*, 65(4), 14-18.
- Chappuis, S., Chappuis, J. & Stiggins, R. (2009). The quest for quality. *Educational Leadership*, 67(3), 14-19.
- Chappuis, S., & Stiggins, R. (2002). Classroom assessment for learning. *Educational Leadership*, 60(1), 40-43.
- Chappuis, J., Stiggins, R., Chappuis, S., & Arter, J. (2012). *Classroom assessment for student learning: Doing it right—Using it well* (2nd ed.). Portland, OR: Pearson Assessment Training Institute.
- Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE (American Association for Higher Education and Accreditation) Bulletin*, 39(7), 3–7. Recuperado de <http://www.aahea.org/articles/sevenprinciples1987.htm>
- Chickering, A. W., & Gamson, Z. F. (1999). Development and adaptations of the seven principles for good practice in undergraduate education. *New Directions for Teaching & Learning*, 80, 75-81. doi: 10.1002/tl.8006
- Chocarro, E., Sobrino, A., & González, M. C. (2013). Scholarship of teaching and learning: Un modelo de desarrollo profesional de los profesores universitarios. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 16(1), 5-14. doi: 10.6018/reifop.16.1.179401
- Cid, A. P. (2012). Concordancia vs. discordancia en la evaluación del practicum: Protocolos, instrumentos y elementos personales del proceso. *Revista de Docencia*

Universitaria, 10(2), 381-394. Recuperado de http://red-u.net/redu/documentos/vol10_n2_completo.pdf

Cid-Sabucedo, A., Pérez-Abellás, A., & Zabalza, M. A. (2009). Las prácticas de enseñanza declaradas de los 'mejores profesores' de la Universidad de Vigo. *Revista Electrónica de Investigación y Evaluación Educativa*, 15(2). Recuperado de http://www.uv.es/relieve/v15n2/RELIEVEv15n2_7.pdf

Cisneros-Cohernour, E., & Stake, R. (2010). La evaluación de la docencia en educación superior: De evaluaciones basadas en opiniones de estudiantes a modelos por competencias. *Revista Iberoamericana de Evaluación Educativa*, 3(1e), 218-231. Recuperado de http://www.rinace.net/riee/numeros/vol3-num1_e/art17.pdf

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24(2), 205-249. doi: 10.1007/s10648-011-9191-6

Clipa, O., & Duta, N. (2010). La percepción del alumnado acerca de la frecuencia de la evaluación. *Revista d'Innovació i Recerca en Educació*, 3(2), 53-67. Recuperado de <http://www.raco.cat/index.php/REIRE/article/view/196165/262999>

Coe, R. (2000). *What is an effect size: A guide for users*. Recuperado de www.cemcentre.org/ebeuk/research/effectsize/ESguide

Coiduras, J. (2009). Escuela y universidad, contextos en alternancia para la formación de docentes. Retos y propuestas ante las nuevas titulaciones de grado. *Organización y Gestión Educativa*, 3, 10-13.

Coll, C. (1983). La evaluación en el proceso de enseñanza-aprendizaje. *Cuadernos de Pedagogía*, 103-104, 13-17.

Coll, C., Barberá, E., & Onrubia, J. (2000). La atención a la diversidad en las prácticas de evaluación. *Infancia y Aprendizaje*, 90, 111-132.

- Coll, C., & Onrubia, J. (1999). Evaluación de los aprendizajes y atención a la diversidad. En C. Coll (Coord.), *Psicología de la Instrucción: La enseñanza y el aprendizaje en la educación secundaria* (pp. 141-168). Barcelona: Horsori-ICE UB.
- Coll, C., & Onrubia, J. (2002). Evaluar en una escuela para todos. *Cuadernos de Pedagogía*, 318, 50-54.
- Coll, C., Rochera, M. J., Mayordomo, R. M., & Naranjo, M. (2007). Continuous assessment and support for learning: An experience in educational innovation with ICT support in higher education. *Electronic Journal of Research in Educational Psychology*, 13(5), 783-804. Recuperado de http://www.investigacion-psicopedagogica.org/revista/articulos/13/english/Art_13_205.pdf
- Coll, C., Rochera, M. J., & Onrubia, J. (2009). De la evaluación continua hacia la autorregulación del aprendizaje. Algunos criterios y propuestas en la enseñanza superior. En M. Castelló (Coord.), *La evaluación auténtica en enseñanza secundaria y universitaria: Investigación e innovación* (pp. 117-144). Barcelona: Edebé.
- Comas, M. A. (2013). El EEES, identidad y competitividad europea: Principios fundamentales e interpretación de las principales autoridades. *Revista de Docencia Universitaria*, 11(1), 243-263 Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/380>
- Combs, K. L., Gibson, S. K., Hays, J. M., Saly, J., & Wendt, J. T. (2008). Enhancing curriculum and delivery: Linking assessment to learning objectives. *Assessment & Evaluation in Higher Education*, 33(1), 87-102. doi: 10.1080/02602930601122985
- Comisión de las Comunidades Europeas. (2007). *Libro verde. El espacio europeo de investigación: Nuevas perspectivas*. Recuperado de http://ec.europa.eu/research/era/pdf/era_gp_final_es.pdf

Comisión Europea. (2010). *Europa 2020: Una estrategia para un crecimiento inteligente, sostenible e integrador*. Recuperado de http://ec.europa.eu/commission_2010-2014/president/news/documents/pdf/20100303_1_es.pdf

Comisión Europea. (2013). *Horizon 2020. Council decision establishing the specific programme implementing Horizon 2020 - the framework programme for research and innovation (2014-2020)*. Recuperado de <http://www.eshorizonte2020.es/que-es-horizonte-2020/documentos-de-interes/programa-especifico-de-horizonte-2020>

Comunicado de la Conferencia de Ministros Europeos de Educación en Lovaina. (2009). *El proceso de Bolonia 2020 EEES en la nueva década*. Recuperado de http://www.ond.vlaanderen.be/hogeronderwijs/bologna/conference/documents/Leuven_Louvain-la-Neuve_Communicu%C3%A9 April_2009.pdf

Consejo Europeo extraordinario de Lisboa. (2000). *Hacia la Europa de la innovación y el conocimiento*. Recuperado de http://europa.eu/legislation_summaries/education_training_youth/general_framework/c10241_es.htm

Contreras, L. C. (2007). El marco de la convergencia europea: Los créditos ECTS. En M. C. Fonseca & J. I. Aguaded (Dirs.), *Enseñar en la universidad. Experiencias y propuestas para la docencia universitaria* (pp. 3-12). La Coruña: Netbiblo.

Contreras, J. (2010). Ser y saber en la formación didáctica del profesorado: Una visión personal. *Revista Interuniversitaria de Formación del Profesorado*, 68(24,4), 61-82.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7). Recuperado de <http://pareonline.net/pdf/v10n7.pdf>

Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, 64, 672-683.

Crossouard, B. (2010). Reforms to higher education assessment reporting: Opportunities and challenges. *Teaching in Higher Education*, 35(3), 247-258. doi: 10.1080/13562511003740809

Cruz de la, M., Pozo, J. I., Huarte, M. F., & Scheuer, N. (2006). Concepciones de enseñanza y prácticas discursivas en la formación de futuros profesores. En J. I. Pozo et al. (Eds.), *Nuevas formas de pensar la enseñanza y el aprendizaje. Concepciones de profesores y alumnos* (pp. 359-371). Barcelona: Graó.

D

Da Silva, E. (2007). A ação docente na educação a distância: Competências para a mediação em rede. *Revista Electrónica Teoría de la Educación. Educación y Cultura en la Sociedad de la Información*, 8(2), 69-85. Recuperado de http://www.usal.es/~teoriaeducacion/rev_numero_08_02/n8_02_oliveira.pdf

Darby, J. (2007). Evaluating course evaluations: The need to establish what is being measured. *Assessment & Evaluation in Higher Education*, 32 (4), 441-455. doi: 10.1080/02602930600898577

Darling-Hammond, L., Herman, J., Pellegrino, J., Abedi, J., Aber, J. L., Baker, E., ... Steelle, C. M. (2013). *Criteria for high-quality assessment*. Recuperado del sitio de Internet del Stanford Center for Opportunity Policy in Education (SCOPE): https://edpolicy.stanford.edu/sites/default/files/publications/criteria-higher-quality-assessment_2.pdf

Davenport, T. H. (2006). *Motivar, retener y crear valor en la era del conocimiento*.

Barcelona: Ediciones Deusto.

Delbanco, A. (2012). *College: What it was, is, and should be*. Princeton University Press.

Declaración de Bolonia. (1999). *El espacio Europeo de Educación Superior. Declaración conjunta de los Ministros Europeos de Educación*. Recuperado de

http://www.educacion.gob.es/boloniaensecundaria/img/Declaracion_Bolonia.pdf

Declaración de La Sorbona. (1988). *Declaración conjunta para la armonización del diseño del sistema de educación superior europeo*. Recuperado de

<http://tecnologiaedu.us.es/mec2011/htm/mas/2/21/15.pdf>

Comunicado de la Conferencia de Ministros Europeos de Educación en Lovaina. (2009). *El proceso de Bolonia 2020-EEES en la nueva década*. Recuperado de

http://www.ond.vlaanderen.be/hogeronderwijs/bologna/conference/documents/Leuven_Louvain-la-Neuve_Communicu%C3%A9 April_2009.pdf

Denzin, N. K., & Lincoln, Y. S. (Coords.). (2012). *Manual de investigación cualitativa* (Vol II).

Barcelona: Gedisa.

Díaz, A. (2006). El enfoque de competencias en la educación, ¿alternativa o disfraz de cambio? *Perfiles Educativos*, 28(11), 7-36. Recuperado de

<http://www.scielo.org.mx/pdf/peredu/v28n111/n111a2.pdf>

Dixon, R., & Rawling, G. (1987). Experiences with continuous assessment. *Assessment & Evaluation in Higher Education*, 12(1), 24-36. doi: 10.1080/0260293870120103

Dochy, F., Segers, M., & Dierick, S. (2002). Nuevas vías de aprendizaje y enseñanza y sus consecuencias: Una era de evaluación. *Revista de Docencia Universitaria*, 2(2).

Recuperado de <http://red->

[u.net/redu/documentos/volumenes/vol2_n2/vol2_n2.htm#dochy](http://red-u.net/redu/documentos/volumenes/vol2_n2/vol2_n2.htm#dochy)

- Dochy, F., Segers, M., & Sluijmans, D. (1999). The use of self-, peer-, and co-assessment in higher education: A review. *Studies in Higher education*, 24(3), 331-350. doi: 10.1080/03075079912331379935
- Duncan, N. 2007. Feedforward: Improving students' use of tutors' comments. *Assessment & Evaluation in Higher Education*, 32(3), 271-283. doi: 10.1080/02602930600896498
- Durán, D. (2009). GAPPISA, una guía para el análisis de pruebas evaluativas desde la perspectiva PISA. En C. Monereo (Coord.), *Pisa como excusa. Repensar la evaluación para cambiar la enseñanza* (pp. 111-124). Barcelona: Graó.

E

- Edwards, P. A., Turner, J. D., & Mokhtari, K. (2008). Balancing the assessment of learning and for learning in support of student literacy achievement. *Reading Teacher*, 61(8), 682-684. doi:10.1598/RT.61.8.12
- Elexpuru, I., Martínez, A., Villardón, L., & Yániz, C. (2006). Diseño, desarrollo y evaluación del plan de formación y desarrollo del profesorado para la incorporación del Modelo de Formación de la Universidad de Deusto (Cuadernos monográficos del ICE, nº 13). Bilbao: Universidad de Deusto.
- Elliot, J. (1986). Autoevaluación, desarrollo profesional y responsabilidad. En M. Galton & B. Moon (Dirs.), *Cambiar la escuela, cambiar el currículum* (pp. 237-259). Barcelona: Martínez Roca.
- Engel, A., & Bustos, A. (2009). Evaluación de la competencia digital en primaria y secundaria. *Aula de Innovación Educativa*, 181, 17-21.

- Entwistle, N. (noviembre, 2000). *Promoting deep learning through teaching and assessment: Conceptual frameworks and educational contexts*. Paper presentado en la Teaching and Learning Research Programme (TLRP) Conference. Leicester. Recuperado de <http://www.etl.tla.ed.ac.uk/docs/entwistle2000.pdf>
- Entwistle, N. (2005). Contrasting perspectives on learning. En F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning. Implications for teaching and studying in higher education* (3rd ed.) (pp. 3-22). Recuperado de http://www.docs.hss.ed.ac.uk/iad/Learning_teaching/Academic_teaching/Resources/Experience_of_learning/EoLChapter1.pdf (Trabajo original publicado en 1984).
- Erwin, D. (2003). Evaluación y valoración: Una aproximación sistemática. En S. Brown & A. Glasner (Eds.), *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 49-60). Madrid: Narcea.
- Escamilla, A. (2010). Evaluación de competencias. En *Primeras jornadas internacionales sobre el EEES: Evaluación* (pp. 19-24). Barcelona: Universidad Internacional de Catalunya.
- Escudero, T. (1987). Buscando una mejor selección de universitarios. *Revista de Educación*, 283, 249-283.
- Escudero, J. M. (2006). La formación del profesorado y la garantía del derecho a una buena educación para todos. En J. M. Escudero & A. Luis (Coords.), *La formación del profesorado y la mejora de la educación para todos: Políticas y prácticas* (pp. 21-51). Barcelona: Octaedro.
- Escudero, J. M. (2008). Las competencias profesionales y la formación universitaria: Posibilidades y riesgos. *Revista de Docencia Universitaria*, 6(2). Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/69>

- Escudero, T. (2003). Desde los test hasta la investigación evaluativa actual. Un siglo, el XX, de intenso desarrollo de la evaluación en educación. *Revista Electrónica de Investigación y Evaluación Educativa*, 9(1). Recuperado de http://www.uv.es/RELIEVE/v9n1/RELIEVEv9n1_1.htm
- Esteve, J. M. (2006). La profesión docente en Europa: Perfil, tendencias y problemática. La formación inicial. *Revista de Educación*, 340, 19-40. Recuperado de http://www.revistaeducacion.mec.es/re340_02.html
- Esteve, O. (2010). La evaluación desde la perspectiva del aprendizaje autorregulado. En *Primeras jornadas internacionales sobre el EEES: Evaluación* (pp. 26-33). Barcelona: Universidad Internacional de Catalunya.
- Esteve, F. M., & Gisbert, M. (2011). El nuevo paradigma del aprendizaje y las nuevas tecnologías. *Revista de Docencia Universitaria*, 9(3), 55-73. Recuperado de http://red-u.net/redu/documentos/volumenes_completos_pdf/vol9_n3_completo.pdf
- Etxague, X. (Coord.). (2004). *Didaktika Orokorra*. Donostia: Erein.
- European Association for Quality Assurance in Higher Education (ENQA). (2009). *Standards and guidelines for quality assurance in the European Higher Education Area* (3rd ed.). Recuperado de http://www.enqa.eu/wp-content/uploads/2013/06/ESG_3edition-2.pdf
- European Higher Education Area (EHEA). (2010). *Bologna Process. European Higher Education Area* [Sitio de Internet oficial]. Recuperado de <http://www.ehea.info/>
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, 83(1), 70-120. doi: 10.3102/0034654312474350

F

- Fairweather, J. S. (2005). Beyond the rhetoric: Trends in the relative value of teaching and research in faculty salaries. *The Journal of Higher Education*, 76(4), 401-422.
Recuperado de <http://www.jstor.org/stable/3838846>
- Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. London: Routledge.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: Metanalysis. *Review of Educational Research*, 59(4), 395-430. doi: 10.3102/00346543059004395
- Falchikov, N., & Boud, D. (2007). Assessment and emotion: The impact of being assessed. En D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term* (pp. 144-156). London: Routledge.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322. doi: 10.3102/00346543070003287
- Feixas, M. (2002). *El desenvolupament professional del professor universitari com a docent* (Tesis doctoral). Recuperado de <http://ddd.uab.cat/record/38084>
- Feixas, M. (2008). El papel del profesor universitario en el EEES. En M. Hijano (Coord.), *Las titulaciones de educación ante el Espacio Europeo de Educación Superior. Análisis de experiencias* (pp. 13-45). Málaga: Ediciones Aljibe.
- Feixas, M., Fernández, A., Lagos, P., Quesada, C., & Sabaté, S. (2013). Factores condicionantes de la transferencia de la formación docente en la universidad: Un estudio sobre la transferencia de las competencias docentes. *Infancia y Aprendizaje*:

Journal for the Study of Education and Development, 36(3), 401-416. doi:
10.1174/021037013807533034

Feldman, K. A. (1988). Effective college teaching from the students' and faculty's view: Matched or mismatched priorities? *Research in Higher Education*, 28(4), 291-344. doi: 10.1007/BF01006402

Ferguson, P. (2011). Student perceptions of quality feedback in teacher education. *Assessment & Evaluation in Higher Education*, 36(1), 51-62. doi:
10.1080/02602930903197883

Fernández, A. (2008). La formación inicial del profesorado universitario: El título de especialista en Pedagogía Universitaria de la Universidad Politécnica de Valencia. *Revista Interuniversitaria de Formación del Profesorado*, 63(22,3), 161-187.

Fernández, A. (2010). La evaluación orientada al aprendizaje en un modelo de formación por competencias en la educación universitaria. *Revista de Docencia Universitaria*, 8(1), 11-34. Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/144>

Fernández, M. (1989). *Así enseña nuestra universidad. Hacia una construcción crítica de una didáctica universitaria*. Universidad Complutense de Madrid.

Fernández, M .L. (2007). Planificación del aprendizaje. En M. C. Fonseca & J. I. Aguaded (Dirs.), *Enseñar en la universidad. Experiencias y propuestas para la docencia universitaria* (pp. 13-24). La Coruña. Netbiblo.

Fernández, I., & Márquez, D. (2014). ¿Formación docente o desarrollo educacional? Situación actual y tendencias emergentes en las universidades del estado español. En C. Monereo (Coord.), *Enseñando a enseñar en la universidad. Formación del profesorado basada en incidentes críticos* (pp. 251-287). Barcelona: Octaedro.

- Flemming, N. D. (2003). Calidad y objetividad en la corrección del trabajo escrito. En S. Brown & A. Glasner (Eds.), *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 105-113). Madrid: Narcea.
- Flórez, J. (1999). *Evaluación de la calidad de la docencia*. Universidad de León.
- Fonseca, M. C. (2010). La innovación docente, una visión de conjunto. *Cuadernos de Pedagogía*, 403, 44-47.
- Font, A. (2010). Exigències i problemes de l'avaluació en l'Espai Europeu d'Educació Superior. *Temps d'Educació*, 38, 51-62. Recuperado de <http://www.raco.cat/index.php/TempsEducacio/article/view/211244>
- Fook, C. Y., & Sidhu, G. K. (2010). Authentic assessment and pedagogical strategies in higher education. *Journal of Social Science*, 6, 153-161. doi: 10.3844/jssp.2010.153.161
- Forés, A., Sánchez, J. A., & Sancho, J. M. (2014). Salir de la zona de confort. Dilemas y desafíos en el EEES. *Tendencias Pedagógicas*, 23, 205-214. Recuperado de http://www.tendenciaspedagogicas.com/Articulos/2014_23_15.pdf
- Fox, D. (1981). *El proceso de investigación en educación*. Pamplona: Eunsa.
- Fraile, A., López, M. V., Castejón, F. J., & Romero, R. (2013). La evaluación formativa en docencia universitaria y el rendimiento académico del alumnado. *Aula Abierta*, 41(2), 23-34. Recuperado de <http://dialnet.unirioja.es/ejemplar/329187>
- Frederiksen, N. (1984). The real test bias. Influences of testing on teaching and learning. *American Psychologist*, 39, 193-202. doi: 10.1037/0003-066X.39.3.193
- Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18, 27-32. doi: 10.3102/0013189X018009027

- Frías, M. D., Pascual, J., & García, J. F. (2000). Tamaño del efecto del tratamiento y significación estadística. *Psicothema*, 12(2), 236-240. Recuperado de <http://www.psicothema.com/pdf/555.pdf>
- Fuente de la, J., Vicente, J. M., Sánchez, F. J., & Berbén, A. B. (2010). Percepción del proceso de enseñanza-aprendizaje y rendimiento académico en diferentes contextos instruccionales de la educación superior. *Psicothema*, 22(4), 806-812. Recuperado de <http://www.psicothema.com/pdf/3805.pdf>
- Fuente de la, J., Sander, P., & Putwain, D. (2013). Relación entre la confianza académica, los enfoques de aprendizaje y el rendimiento académico de estudiantes universitarios: El papel del género. *Revista de Psicodidáctica*, 18(2), 375-393. doi: 10.1387/revPsicodidact.7078
- Fuentes, M. E., & Herrero, J. R. (1999). Evaluación docente: Hacia una fundamentación de la autoevaluación. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 2(1), 353-368. Recuperado de http://www.aufop.com/aufop/uploaded_files/articulos/1224341555.pdf
- Fuertes, M. T., & Balaguer, R. (2012). El TFG como elemento de mejora de la calidad en la evaluación del módulo practicum: Propuesta de la Universidad Internacional de Catalunya (UIC). *Revista de Docencia Universitaria*, 10(2), 329-344. Recuperado de http://red-u.net/redu/documentos/vol10_n2_completo.pdf
- Fullan, M. G. (1991). *The new meaning of educational change*. New York: Teachers College Press.
- Fullan, M. G. (2007). *Las fuerzas del cambio: La continuación*. Madrid: Akal.

G

- Gairín, J. (Coord.). (2009). *Guía para la evaluación de competencias en el área de ciencias sociales*. Recuperado del sitio de Internet de la Agència per a la Qualitat del Sistema Universitari de Catalunya (AQU): http://www.aqu.cat/doc/doc_14646947_1.pdf
- Gairín, J. (2011). Formación de profesores basada en competencias. *Bordón*, 63(1), 93-102.
- Gairín, J., & Darder, P. (Coord.). (1994). *Organización de centros educativos: Aspectos básicos*. Barcelona: Praxis.
- Galán, A., & Rubalcaba, L. (2007). Fines, motivación y ámbito del trabajo en la universidad. En A. Galán (Ed.), *El perfil del profesor universitario* (pp. 33-55). Madrid: Encuentro.
- Gallego, M. J. (2007). Las funciones docentes presenciales y virtuales del profesorado universitario. *Revista Electrónica Teoría de la Educación. Educación y Cultura en la Sociedad de la Información*, 8(2), 137-161. Recuperado de http://www.usal.es/~teoriaeducacion/rev_numero_08_02/n8_02_arrufat.pdf
- García, J. M. (1997). Valoración de la competencia docente del profesor universitario. *Revista Complutense de Educación*, 8(2), 81-100. Recuperado de <http://revistas.ucm.es/index.php/RCED/article/view/RCED9797220081A/17431>
- García, M. J. (2010). *Diseño y validación de un modelo de evaluación por competencias en la universidad* (Tesis doctoral, Universidad Autónoma de Barcelona). Recuperada de <http://www.tesisenred.net/handle/10803/5065>
- García, B., Loredo, J., Luna, E., & Rueda, M. (2008). Modelo de evaluación de competencias docentes para la educación media y superior. *Revista Iberoamericana de Evaluación Educativa*, 1(3e), 124-136. Recuperado de http://www.rinace.net/riee/numeros/vol1-num3_e/art8.pdf

- García, M. P., & Maquilón, J. J. (2010). El futuro de la formación del profesorado universitario. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 14(1), 17-26. Recuperado de http://aufop.com/aufop/uploaded_files/articulos/1301587561.pdf
- García, B., Valencia, A., & Pineda, V. (2012). Diseño y validación de un instrumento para la auto-evaluación de las competencias docentes. *Revista Iberoamericana de Evaluación Educativa*, 5(1), 66-83. Recuperado de http://www.rinace.net/riee/numeros/vol5-num1_e/art6.pdf
- García-Valcárcel, A. (1992). Características del buen profesor universitario según estudiantes y profesores. *Revista de Investigación Educativa*, 19(1), 31-50. Recuperado de <http://revistas.um.es/rie/article/view/136551/124161>
- García-Valcárcel, A. (1993). Análisis de los modelos de enseñanza empleados en el ámbito universitario. *Revista española de pedagogía*, 194, 27-53.
- Gargallo, A. (2009). Aproximación al estudio de la evaluación en la universidad: Un análisis exploratorio con alumnos semipresenciales. *Contextos Educativos*, 12, 77-93. Recuperado de <https://publicaciones.unirioja.es/ojs/index.php/contextos/article/view/612>
- Gargallo, B., Sánchez, F., Ros, C., & Ferreras, A. (2010). Estilos docentes de los profesores universitarios. La percepción de los alumnos de los buenos profesores. *Revista Iberoamericana de Educación*, 51(4). Recuperado de <http://www.rieoei.org/deloslectores/3236Lopez.pdf>
- Gargallo, B., Suárez, J., Garfella, P. R., & Fernández, A. (2011). El cuestionario CEMEDEPU. Un instrumento para la evaluación de la metodología docente y evaluativa de los profesores universitarios. *Estudios Sobre Educación*, 21, 9-40. Recuperado de

http://dadun.unav.edu/bitstream/10171/22511/2/Art%C3%ADculo_1_El%20cuestionario%20CEMEDEPU.pdf

Gedye, S. (2010). Formative assessment and feedback: A review. *Planet*, 23, 40-45. doi: 10.11120/plan.2010.00230040

Gibbs, G. (2003). Uso estratégico de la evaluación en el aprendizaje. En S. Brown & A. Glasner (Eds.), *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 61-74). Madrid: Narcea.

Gibbs, G. (2006). How assessment frames student learning. En C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 23-36). New York: Routledge.

Gibbs, G. (2008). International perspectives on the development of teaching in research universities. En, *Actas del V Congreso Internacional Docencia Universitaria e Innovación: El Cambio en la cultura docente universitaria* [CD]. Universidad de Lleida.

Gibbs, G., (2010). *Dimensions of quality*. Recuperado de http://www.celt.mmu.ac.uk/policy/ltmmu/docs/Dimensions_of_Quality%20%20Graham%20Gibbs.pdf

Gibbs, G., & Dunbar-Goddet, H. (2007). *The effects of programme assessment environments on student learning*. Recuperado del sitio de Internet de la Higher Education Academy: http://www.heacademy.ac.uk/assets/documents/teachingandresearch/gibbs_0506.pdf

Gibbs, G., & Simpson, C. (septiembre, 2003). *Measuring the response of students to assessment: The Assessment Experience Questionnaire*. Trabajo presentado en el 11th International Symposium Improving Students Learning: Theory, Research and

- Scholarship. Hinckley, UK. Recuperado de <http://www.open.ac.uk/fast/pdfs/Gibbs%20and%20Simpson%202003.pdf>
- Gibbs, G., & Simpson, C. (2009). *Condiciones para una evaluación continuada favorecedora del aprendizaje* (Cuadernos de docencia universitaria, 13). Recuperado de <http://www.octaedro.com/ice/pdf/13CUADERNO.pdf>
- Gielen, S., Dochy, F., & Dierick, S. (2003). Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post and true assessment effects. En M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 37-54). doi: 10.1007/0-306-48125-1_3
- Gil-Flores, J. (2012). La evaluación del aprendizaje en la universidad según la experiencia de los estudiantes. *Estudios Sobre Educación*, 22, 133-153. Recuperado de <http://dadun.unav.edu/bitstream/10171/22636/2/ESE22-07-Gil-Flores.pdf>
- Glasner, A. (2003). Innovaciones en la evaluación del estudiante: Un sistema de amplia perspectiva. En S. Brown & A. Glasner (Eds.), *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 23-33). Madrid: Narcea.
- Gómez, M. A., Rodríguez, G., & Ibarra, M. S. (2013). COMPES: Autoinforme sobre las competencias básicas desarrolladas relacionadas con la evaluación de los estudiantes. *Estudios Sobre Educación*, 24, 197-224. Recuperado de <http://dadun.unav.edu/bitstream/10171/29570/2/RODRIGUEZ%20IBARRA.pdf>
- González, I. (2003). Determinación de los elementos que condicionan la calidad de la universidad: Aplicación práctica de un análisis factorial. *Revista Electrónica de Investigación y Evaluación Educativa*, 9(1). Recuperado de http://www.uv.es/RELIEVE/v9n1/RELIEVEv9n1_4.htm

- González, P., & Grande, F. J. (1999). Experiencia en la evaluación de la universidad: El caso del profesorado. *Revista Interuniversitaria de Formación del Profesorado*, 34, 61-67.
- González, I., & López, A. (2010). Sentando las bases para la construcción de un modelo de evaluación a las competencias docentes del profesorado universitario. *Revista de Investigación Educativa*, 28 (2), 403-423. Recuperado de <http://revistas.um.es/rie/article/view/109431/121661>
- González, M., & Raposo, M. (2009). Valoraciones del profesorado universitario sobre las acciones que pueden favorecer el proceso de convergencia europea. *Revista de Educación*, 349, 361-390. Recuperado de http://www.revistaeducacion.mec.es/re349/re349_17.pdf
- González, J., & Wagenaar, R. (Eds.). (2003). *Tuning educational structures in Europe. Informe final*. Bilbao. Universidad de Deusto- Universidad de Groningen.
- Goñi, J. M. (2005). *El espacio europeo de educación superior, un reto para la universidad. Competencias, tareas y evaluación, los ejes del currículum universitario*. Barcelona: Octaedro.
- Goodwin, L. D., & Leech, N. L. (2003). The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. *Measurement and Evaluation in Counseling and Development*, 33, 181-191.
- Gow, L., & Kember, D. (1993). Conceptions of teaching and their relationship to student learning. *British Journal of Educational Psychology*, 63, 20-33. doi: 10.1111/j.2044-8279.1993.tb01039.x

- Gros, B. (2006). La investigación sobre docencia universitaria: Retos e iniciativas. En M. Martínez & S. Carrasco (Coords.), *Propuestas para el cambio docente en la universidad* (pp. 33-48). Barcelona: Octaedro-ICE UB.
- Gros, B., & Romañà, T. (2004). *Ser profesor. Palabras sobre la docencia universitaria* (ed. rev.). Universitat de Barcelona.
- GRup d'Avaluación de la Práctica Académica (GRAPA), Recerca i Innovació en Metodologies de l'Aprenentatge de la Universidad Politécnica de Catalunya RIMA (RIMA). (julio, 2012). *Criterios para prácticas de evaluación de calidad. Una propuesta de GRAPA-RIMA*. Trabajo presentado en el VII Congreso Internacional en Docencia Universitaria e Innovación. Barcelona. Recuperado de: www.cidui.org/revista-cidui12/index.php/cidui12/article/.../69/59
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Guenaga, M., Arranz, S., Bezanilla, M. J., Rubio, I., Ortiz de Guinea, A., ... Menchaca, I. (2013). Serious games for the development of employment oriented competentes. *Revista Iberoamericana de Tecnologías del Aprendizaje*, 8(4), 176-183. doi: 10.1109/RITA.2013.2285024
- Guerrero, A., Huertas, M., Mor, E., & Rodríguez, M. (2013). Explicitando la interrelación entre las actividades de aprendizaje, el proceso de evaluación y la adquisición de competencias. *Revista Iberoamericana de Educación a Distancia*, 16(2), 127-146. doi: 10.5944/ried.16.2.9909
- Gulikers, J. T., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology, Research and Development*, 32(3), 67-86. doi: 10.1007/BF02504676

- Gullickson, A. (2012). Matching measurement instruction to classroom-based evaluation: Perceived discrepancies, needs and challenges. En S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 1-26). Recuperado de <http://digitalcommons.unl.edu/burosteachertraining/> (Trabajo original publicado en 1993).
- Gutiérrez, C., Pérez, A., Pérez, M., & Palacios, A. (2011). Percepciones de profesores y alumnos sobre la enseñanza, evaluación y desarrollo de competencias en estudios universitarios de formación de profesorado. *Cultura y Educación*, 23(4), 499-514. doi: 10.1174/113564011798392451
- Gutiérrez, C, Pérez, A., & Pérez, M. (2013). Percepciones de profesores, alumnos y egresados sobre los sistemas de evaluación en estudios universitarios de formación del profesorado de educación física. *Ágora para la Educación Física y el Deporte*, 15(2), 130-151. Recuperado de http://agora-revista.blogs.uva.es/files/2013/11/agora_15_2d_gutierrez_et_al.pdf

H

- Hafner, J., & Hafner, P. (2003). Quantitative analysis of the rubric as an evaluation tool: An empirical study of student peer-group rating. *International Journal of Science Education*, 25(12), 1509-1528. doi: 10.1080/0950069022000038268
- Hager, P., & Gonczi, A. (1994). General issues about assessment of competence. *Assessment & Evaluation in Higher Education*, 19(1), 3-14. doi: 10.1080/0260293940190101

- Haigh, M. (2007). Sustaining learning through assessment: An evaluation of the value of a weekly class quiz. *Assessment & Evaluation in Higher Education*, 32(4), 457-474. doi: 10.1080/02602930600898593
- Hamodi, C., & López, T. (2012). La evaluación formativa desde la perspectiva del alumnado y egresados. *Psychology, Society, & Education*, 4(1), 103-116. Recuperado de dialnet.unirioja.es/descarga/articulo/3961366.pdf
- Hannan, A., & Silver, H. (2005). *La innovación en la educación superior*. Madrid. Narcea.
- Hargreaves, E. (2007). The validity of collaborative assessment for learning. *Assessment in Education*, 14(2), 185-199. doi: 10.1080/09695940701478594
- Harlen, W. (Dir.). (2003). *The role of teachers in the assessment of learning* (Proyecto Assessment Systems for de Future, Assessment Reform Group). Recuperado del sitio de Internet de la Nuffield Foundation: http://www.nuffieldfoundation.org/sites/default/files/assessment_booklet.pdf
- Harlen, W. (2005). Trusting teachers' judgments: Research evidence of the reliability and validity of teachers' assessment for summative purposes. *Research Papers in Education*, 20(3), 245-270. doi: 10.1080/02671520500193744
- Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. En J. Gardner (Ed.), *Assessment and learning* (pp. 103-118). London: Sage.
- Harlen, W. (2007). Criteria for evaluating systems for student assessment. *Studies in Educational Evaluation*, 33, 15-28. doi: 10.1016/j.stueduc.2007.01.003
- Harlen, W., & Deakin, R. (2003). Testing and motivation for learning. *Assessment in Education*, 10(2), 169-207. doi: 10.1080/0969594032000121270

- Harrington, K., Elander, J., Lusher, J. Norton, L. Aiyegbayo, O., Pitt, ... Reddy, P. (2006). *Using core assessment criteria to improve essay writing*. En C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education*. (pp. 110-119). New York: Routledge.
- Hartley, J., & Chesworth, K. (2000). Qualitative and quantitative methods in research of essay writing: No one way. *Journal of Further and Higher Education*, 24, 15-24. doi: 10.1080/030987700112282
- Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills intervention on student learning: A meta-analysis. *International Journal of Educational Research*, 11, 187-212. doi: 10.3102/00346543066002099
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. doi: 10.3102/003465430298487
- Heathfield, M. (2003). El aprendizaje en grupo para fomentar un aprendizaje de calidad. En S. Brown & A. Glasner (Eds.), *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 155-165). Madrid: Narcea.
- Herman, J. L., Aschbacher, P., & Winters, L. (1992). *A practical guide to alternative assessment*. California: Association for Supervision and Curriculum Development (ASCD).
- Hernández, F. (1996). La evaluación de los alumnos en el contexto de la evaluación de la calidad de las universidades. *Revista de Investigación Educativa*, 14(2), 25-50. Recuperado de <http://revistas.um.es/rie/article/view/141222/127072>
- Hernández, F. (2002). Docencia e investigación en la educación superior. *Revista de Investigación Educativa*, 20(2), 271-301. Recuperado de <http://revistas.um.es/rie/article/view/98921>

- Hernández, F. (2005). *Aprendizaje, competencias y rendimiento en educación superior*. Madrid: La Muralla.
- Hernández, F. (2014). Evaluación y acreditación del profesorado, programas e instituciones educativas. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 17(1), 15-32. doi: 10.6018/reifop.17.1.198821
- Hernández, F., García, M. P., & Maquilón, J. J. (2001). Estudio empírico de los enfoques de aprendizaje de los estudiantes universitarios en función del perfil de su titulación (profundo vs. superficial). *Revista Española de Orientación Psicopedagógica*, 12(22), 303-318. Recuperado de <http://www.uned.es/reop/pdfs/2001/12-22-2---303-Fuentsanta%20Hernandez%20Pina.PDF>
- Hernández, F., Martínez, P., Martínez, M., & Monroy, F. (2009). Aprendizaje y competencias. Una nueva mirada. *Revista Española de Orientación Psicopedagógica*, 20(3), 312-319. Recuperado de <http://www.uned.es/reop/pdfs/2009/20-3%20-%20Fuentsanta%20Hernandez%20Pina.pdf>
- Hernández, F., & Maquilón, J. (2011). Las creencias y las concepciones. Perspectivas complementarias. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 14(1), 165-175. Recuperado de <http://www.aufop.com/aufop/revistas/arta/digital/158/1627>
- Hernández, F., Maquilón, J., García, M. P., & Monroy, F. (2010). Concepciones de la enseñanza y el aprendizaje en profesorado de educación superior. *Psicología Educativa*, 16(2), 95-105. doi: 10.5093/ed2010v16n2a1
- Hernández, F., Rosário, P., Cuesta, J. D., Martínez, P., & Ruiz, E. (2006). Promoción del aprendizaje estratégico en estudiantes de primero de universidad: Evaluación de una

intervención. *Revista de Investigación Educativa*, 24(2), 615-632. Recuperado de <http://revistas.um.es/rie/article/view/97221>

Hernández, F., Sales, P. J., & Cuesta, J. D. (2010). Impacto de un programa de autorregulación del aprendizaje en estudiantes de grado. *Revista de Educación*, 353, 571-588. Recuperado de <http://www.mecd.gob.es/dctm/revista-de-educacion/articulosre353/re35321.pdf?documentId=0901e72b812048c2>

Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in Higher Education*, 27(1), 53-64. doi: 10.1080/03075070120099368

Higher Education Founcil Council for England (HEFCE). (2014). *UK review of the provision of information about higher education. National Student Survey results and trends analysis 2005-2013*. Recuperado de http://www.hefce.ac.uk/media/hefce/content/pubs/2014/201413/HEFCE2014_13.pdf

Hills, L., & Glover, C. (febrero, 2005). *Evaluating the evaluation tools: Methodological issues in the FAST Project*. Paper presentado en la Bera Conference. Cardiff. Recuperado de <http://www.open.ac.uk/fast/pdfs/Hills%20%20Glover%20BERA%20paper.pdf>

Holmes, N. (2015). Student perceptions of their learning and engagement in response to the use of a continuous e-assessment in an undergraduate module. *Assessment & Evaluation in Higher Education*, 40(1), 1-14. doi:10.1080/02602938.2014.881978

Hortigüela, D., Pérez, A., & Abella, V. (2015). Perspectiva del alumnado sobre la evaluación tradicional y la evaluación formativa. Contraste de grupos en las mismas asignaturas. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 13(1), 35-48. Recuperado de <http://www.rinace.net/reice/numeros/arts/vol13num1/art3.pdf>

- Hounsell, D. (2003). Student feedback, learning and development. En M. Slowey & D. Watson (Eds.), *Higher education and the lifecourse* (pp. 67-78). Milton Keynes: Open University Press.
- Hounsell, D. (2007). Towards more sustainable feedback to students. En D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education. Learning for the longer term* (pp. 101-113). London: Routledge.
- House, E. (1980). *Evaluating with validity*. Beverly Hills: Sage Publications.
- House, E. (1993). *Profesional evaluation. Social impact and political consequences*. London: Sage.
- Hoyle, E. (1986). *The politics of school management*. London: Hodder and Stoughton.
- Huxham, M. (2007). Fast and effective feedback: Are model answers the answer? *Assessment & Evaluation in Higher Education*, 32(6), 601-611. doi: 10.1080/02602930601116946

I

- Ibarra, M. S., & Rodríguez, G. (2010). Aproximación al discurso dominante sobre la evaluación del aprendizaje en la universidad. *Revista de Educación*, 351, 385-407. Recuperado de http://www.revistaeducacion.educacion.es/re351/re351_16.pdf
- Ibarra, M. S., Rodríguez, G., & Gómez, M. A. (2012). La evaluación entre iguales: Beneficios y estrategias para su práctica en la universidad. *Revista de Educación*, 359, 206-231. doi: 10.4438/1988-592X-RE-2011-359-092

Iglesias, S., & Sulé, M. A. (2005). Introducción al análisis multivariable. En J. P. Lèvy & J. Varela (Dirs.), *Análisis multivariable para las ciencias sociales* (pp. 1-41). Madrid: Pearson Prentice.

Interstate New Teachers Assessment and Support Consortium (INTASC). (2011). *Model core teaching standards: A resource for state dialogue*. Recuperado de http://www.ccsso.org/Documents/2011/InTASC_Model_Core_Teaching_Standards_2011.pdf

Ingvarson, L., & Kleinhenz, E. (2006). Estándares profesionales de práctica y su importancia para la enseñanza. *Revista de Educación*, 340, 265-295.

Inzunza, J. (2008). Evaluación, carrera y vocación docente. *Revista Iberoamericana de Evaluación Educativa*, 1(2), 203-207. Recuperado de www.rinace.net/riee/numeros/vol1-num2/art13.pdf

Ion, G., & Cano, E. (2011). El proceso de implantación de la evaluación por competencias en la educación superior. Un estudio sobre el rol de los cargos académicos. *Revista de Investigación en Educación*, 9(2), 248-258. Recuperado de <http://webs.uvigo.es/reined/ojs/index.php/reined/article/viewFile/304/172>

Iranzo, E., Gimeno, E., & Sánchez, P. (2012). Estrategias de evaluación de competencias de los alumnos universitarios a través del trabajo grupal. *Revista Iberoamericana de Evaluación Educativa*, 5(1), 216-229. Recuperado de http://www.rinace.net/riee/numeros/vol5-num1_e/art15.pdf

Isaeva, T. E. (2007). Classification of the collage instructor's professional and personal competencies. *Russian Education and Society*, 49(9), 31-41. doi: 10.2753/RES1060-9393490904

J

- Jackson, M. (2006). 'Serving time': The relationship of good and bad teaching. *Quality Assurance in Education*, 14(4), 385-397. doi: 10.1108/09684880610703965
- Jessop, T. (Dir.). (2012). TESTA. *Transforming the Experience of Students Through Assessment* (2009 – 2012). Recuperado del sitio de Internet de la Higher Education Academy (HEA):
http://www.heacademy.ac.uk/assets/documents/ntfs/projects/Winchester_final_report.pdf
- Jessop, T. (julio, 2013). *Improving student learning from assessment and feedback: A programme-level view*. Trabajo presentado en la University of Warwick SCAP Conference. Warwick, UK. Recuperado de
<http://www.kcl.ac.uk/study/learningteaching/kli/conferences/scap/2013/scap2013-Jessop.pdf>
- Jessop, T., El Hakim, Y., & Gibbs, G. (2014). The whole is greater than the sum of its parts: A large-scale study of students' learning in response to different programme assessment patterns. *Assessment & Evaluation in Higher Education*, 39(1), 73-88. doi: 10.1080/02602938.2013.792108
- Jessop, T., & Maleckar, B. (2014). The influence of disciplinary assessment patterns on student learning: A comparative study. *Studies in Higher Education*. doi: 10.1080/03075079.2014.943170
- Johnston, S. (1996). What can we learn about teaching from our best university teachers? *Teaching in Higher Education*, 1(2), 213-225. doi: 10.1080/1356251960010204

Joint Committee on Standards for Educational Evaluation (JCSEE). (1981). *Standards for evaluations of educational programs, projects and materials*. Nueva York: McGraw-Hill.

Joint Committee on Standards for Educational Evaluation (JCSEE). (1998). *Estándares para la evaluación de programas*. Bilbao: Ediciones Mensajero.

Joint Committee on Standards for Educational Evaluation (JCSEE). (2001). *Estándares de evaluación de personal: Cómo evaluar sistemas de evaluación de educadores*. Bilbao: Ediciones Mensajero.

Joint Committee on Standards for Educational Evaluation (JCSEE). (2007). *Estándares para la evaluación de los aprendizajes de los estudiantes: Cómo valorar las evaluaciones de los aprendizajes de los estudiantes*. Bilbao: Ediciones Mensajero.

Joint Information Systems Committee (JISC). (2010). *Effective assessment in a digital age*. Recuperado de http://www.jisc.ac.uk/media/documents/programmes/elearning/digiassass_eada.pdf

Jonsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education*, 14(1), 63-76. doi: 10.1177/1469787412467125

Jorba, J., & Casellas, E. (Eds.). (1997). *La regulación y la autorregulación de los aprendizajes. Estrategias y técnicas para la gestión social del aula* (Vol. 1). Madrid: Síntesis.

Jorba, J., & Sanmartí, N. (2000). La función pedagógica de la evaluación. En M. Ballester et al. (Eds.), *Evaluación como ayuda al aprendizaje* (pp. 21-44). Barcelona: Graó.

Joughin, G. (2003). Dimensiones y enfoques de la evaluación oral. En S. Brown & A. Glasner (Eds.), *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 166-175). Madrid: Narcea.

- Juan de, J., Pérez, R. M., Gómez-Torres, M. J., Vizcaya, M. F., & Mora, J. M. (2007). Buenas prácticas en la evaluación de la docencia y del profesorado universitario. *Redes de investigación docente: Espacio Europeo de Educación Superior*, 1(2) 155-182. Recuperado de <http://tecnologiaedu.us.es/mec2011/htm/mas/5/51/35.pdf>
- Juanas de, A. (2011). Aproximaciones teóricas sobre las competencias del profesorado de educación superior. *Tendencias Pedagógicas*, 18, 233-250. Recuperado de http://www.tendenciaspedagogicas.com/Articulos/2011_18_12.pdf

K

- Kaneko, M. (septiembre, 2008). *Beyond the politics of competent. Balancing the social claim and the core of higher education*. Paper presentado en el Institutional Management in Higher Education (IMHE). Paris. Recuperado de <http://www.oecd.org/dataoecd/4/12/41217871.pdf>
- Kember, D. (1997). A reconceptualisation of the research into university academic's conceptions of teaching. *Learning and Instruction*, 7(3), 255-275. doi: 10.1016/S0959-4752(96)00028-X
- Kember, D., & Gow, L. (1994). Orientations to teaching and their effect on the quality of student learning. *The Journal of Higher Education*, 65(1), 58-74. doi: 10.2307/2943877
- Kember, D., & Kwan, K. (2000). Lecturers' approaches to teaching and their relationships to conceptions of good teaching. *Learning and Instruction*, 7, 255-275. doi: 10.1023/A:1026569608656
- Kerlinger, F. H. (2002). *Investigación del comportamiento* (4ª ed.). Mexico: McGraw-Hill.

Kindelán, M. P. (2013). Una perspectiva sobre el binomio enseñanza-investigación en la universidad del siglo XXI. *Revista Complutense de Educación*, 24(1), 27-45. doi: 10.5209/rev_RCED.2013.v24.n1.41190

Knight, P. (2005). *El profesorado de educación superior. Formación para la excelencia*. Madrid: Narcea.

Korthagen, F. (2010). La práctica, la teoría y la persona en la formación del profesorado. *Revista Interuniversitaria de Formación del Profesorado*, 68(24,4), 83-99.

Kreber, C. (2005). Reflection on teaching and the scholarship of teaching. *Higher Education*, 50(2), 323-359. doi: 10.1007/s10734-004-6360-2

L

Land, R. (1997). Moving up to complex evaluation systems. *Nacional Center for Research on Evaluation, Standards, & Student Testing*, 7(1), 1-23. Recuperado de http://www.cse.ucla.edu/products/evaluation/cresst_ec1997_3.pdf

Lapaham, A., & Webster, R. (2003). Evaluación realizada por los compañeros: Motivaciones, reflexiones y perspectivas de futuro. En S. Brown & A. Glasner (Eds.), *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 203-210). Madrid: Narcea.

Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 37, 822-832. doi: 10.1037/0022-3514.37.6.822

Latorre, A., Rincón, D., & Arnal, J. (1996). *Bases metodológicas de la investigación educativa*. Barcelona: GR92.

- Laurillard, D. (2002). *Rethinking university teaching: A framework for the effective use of learning technologies* (2nd ed.). London: Routledge Falmer.
- Le Boterf, G. (2001). *Ingeniería de las competencias*. Barcelona: Gestión 2000.
- Leal, E. (2004). Constate la validez de las pruebas. En L. M. Villar (Coord.), *Programa para la mejora de la docencia universitaria* (pp. 543-556). Madrid: Pearson-Prentice Hall.
- León, O., & Montero, I. (2012). *Diseño de investigaciones* (3ª ed. rev.). Madrid: McGraw-Hill.
- Letelier, M. E. (2009). La evaluación de aprendizajes de personas jóvenes y adultas. *Revista Iberoamericana de Evaluación Educativa*, 2(1), 148-161. Recuperado de www.rinace.net/riee/numeros/vol2-num1/art8.pdf
- Ley 1/2013, de 10 de octubre, de Aprendizaje a lo Largo de la Vida.
- Ley Orgánica 6/2001 del 21 de diciembre, de Universidades (LOU).
- Ley Orgánica 4/2007, de 12 de abril, por la que se modifica la Ley Orgánica 6/2001, de 21 de diciembre, de Universidades (LOMLOU).
- Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th. ed.). Hillslade (NJ): Prentice Hall.
- Litwin, E. (1998). La evaluación campo de controversias y paradojas o un nuevo lugar para la buena enseñanza. En A. Camilloni et al. (Eds.), *La evaluación de los aprendizajes en el debate didáctico contemporáneo* (pp. 11-34). Buenos Aires: Paidós Educador.
- Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, 11(3), 279-290. doi: 10.1080/13562510600680582

Ljungman, A., & Silen, C. (2008). Examination involving students as peer examiners.

Assessment & Evaluation in Higher Education, 33(3), 289-300. doi:

10.1080/02602930701293306

López, R. (2001). *Creencias del profesorado universitario sobre evaluación* (Tesis doctoral).

Recuperada de <http://hdl.handle.net/10481/4570>

López, V. M. (2004). Evaluación, calificación, credencialismo y formación inicial del

profesorado: Efectos y patologías generadas en la enseñanza universitaria. *Revista*

Interuniversitaria de Formación del Profesorado, 18(3), 221-232.

López, V. M. (2006). El papel de la evaluación formativa en el proceso de convergencia

hacia el EEES. Análisis del estado de la cuestión y presentación de un sistema de

intervención. *Revista Interuniversitaria de Formación del Profesorado*, 20(3), 93-119.

López, V. M. (2008). Developing formative and shared assessment systems in higher

education. Analysis of results of putting them into practice in early teacher education.

European Journal of Teacher Education, 31(3), 293-311. doi:

10.1080/02619760802208452

López, V. M. (2009a). Fundamentación teórica y revisión del estado de la cuestión. En V. M.

López (Coord.), *Evaluación formativa y compartida en educación superior*.

Propuestas, técnicas, instrumentos y experiencias (pp. 45-64). Madrid: Narcea.

López, V. M. (2009b). Una propuesta genérica de evaluación formativa y compartida para la

docencia universitaria. En V. M. López (Coord.), *Evaluación formativa y compartida*

en educación superior. Propuestas, técnicas, instrumentos y experiencias (pp. 93-

104). Madrid. Narcea.

López, V. M. (2011). El papel de la evaluación formativa en la evaluación por competencias:

Aportaciones de la Red de Evaluación Formativa y Compartida en docencia

- universitaria. *Revista de Docencia Universitaria*, 9(1), 159-173. Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/198/pdf>
- López, V. M. (2012). Evaluación formativa y compartida en la universidad: Clarificación de conceptos y propuestas de intervención desde la Red Interuniversitaria de Evaluación Formativa. *Psychology, Society, & Education*, 4(1), 117-130. Recuperado de <http://dialnet.unirioja.es/servlet/articulo?codigo=3961371>
- López, V. M., Castejón, F. J., & Pérez, A. (2012). ¿Implicar al alumnado en la evaluación en la formación inicial del profesorado? Un estudio de caso de evaluación entre iguales. *Multidisciplinary Journal of Educational Research*, 2(2), 177-201. doi: 10.4471/remie.2012.09
- López, V. M., Castejón, J., Sicilia, A., Navarro, V., & Webb, G. (2011). The process of creating a cross-university network for formative and shared assessment in higher education in Spain and its potential applications. *Innovations in Education and Teaching International*, 48(1), 79-90. doi: 10.1080/14703297.2010.543768
- López, V. M., Fernández-Balboa, J. M., Santos, M. L., & Fraile, A. (2012). Students' self-grading, professor's grading and negotiated final grading at three university programmes: analysis of reliability and grade difference ranges and tendencies. *Assessment & Evaluation in Higher Education*, 37(4), 453-464. doi: 10.1080/02602938.2010.545868
- López, A. B., González, I., & de León, C. (2014). Perfil de un buen docente. Aplicación de un protocolo de evaluación de las competencias del profesorado universitario. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 17(1), 133-148. doi: 10.6018/reifop.17.1.190531
- López, V. M., Manrique, J. C., & Vallés, C. (2011). La evaluación y la calificación en los nuevos títulos de grado. Especial incidencia en la formación inicial del profesorado.

Revista Electrónica Interuniversitaria de Formación del Profesorado, 14(4), 57-72.

Recuperado de http://www.aufop.com/aufop/uploaded_files/articulos/1327436320.pdf

López, V. M., Martínez, L. F., & Julián, J. A. (2007). La red de evaluación formativa, docencia universitaria y Espacio Europeo de Educación Superior (EEES). Presentación del proyecto, grado de desarrollo y primeros resultados. *Revista de Docencia Universitaria*, 5(2). Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/52>

López, V. M., & Palacios, A. (2012). Percepción de los futuros docentes sobre los sistemas de evaluación de su aprendizaje. *Teoría de la Educación. Educación y Cultura en la Sociedad de la Información*, 13(3), 317-341. Recuperado de http://campus.usal.es/~revistas_trabajo/index.php/revistatesi/article/view/9143/9376

Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science*, 40(2), 257-275. doi: 10.1007/s11251-011-9177-2

M

MacLellan, E. (2001). Assessment for learning: The differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education*, 26(4), 307-318. doi: 10.1080/02602930120063466

Madaus, G. F., & Stufflebeam, D. (1984). Educational evaluation and accountability. A review of quality assurance efforts. *American Behavioral Scientist*, 27(5), 649-672. doi: 10.1177/000276484027005007

- Manzano, V., & Braña, T. (2005). Análisis de datos y técnicas de muestreo. En J. P. Lèvy & J. Varela (Dirs.), *Análisis multivariable para las ciencias sociales* (pp. 91-143). Madrid: Pearson Prentice.
- Marbán, J. M. (Coord.). (2012). *Análisis de las herramientas de evaluación de la calidad docente mediante contrastes basados en estándares internacionales de excelencia* [Proyecto EA2011-0113]. Datos duros inéditos.
- Margalef, L. (2009). Aprender a evaluar: Transformar nuestras prácticas universitarias. En M. Castelló (Coord.), *La evaluación auténtica en enseñanza secundaria y universitaria: Investigación e innovación* (pp. 87-111). Barcelona: Edebé.
- Margalef, L., & Álvarez, J. M. (2005). La formación del profesorado universitario para la innovación en el marco de la integración del Espacio Europeo de Educación Superior. *Revista de Educación*, 337, 51-70.
- Marín, E., & Cervi, J. (2006). Modelos de formación docente para el cambio de concepciones en los profesores. En J. I. Pozo et al. (Eds.), *Nuevas formas de pensar la enseñanza y el aprendizaje. Concepciones de profesores y alumnos* (pp. 418-434). Barcelona: Graó.
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76(5), 704-754. doi: 10.1037/0022-0663.76.5.707
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-388. doi: 10.1016/0883-0355(87)90001-2
- Marso, R. N., & Pigge, F. L. (2012). Teacher training in measurement and assessment skills. En S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp.

129-185). Recuperado de <http://digitalcommons.unl.edu/burosteachertraining/>

(Trabajo original publicado en 1993).

Martilla J. A., & James J. C. (1977). Importance-performance analysis. *Journal of Marketing*, 41(1), 77-79. doi: 10.2307/1250495

Martín, E. (2003). Conclusiones: Un currículo para desarrollar la autonomía del estudiante. En C. Monereo & J. I. Pozo (Eds.), *La universidad ante la nueva cultura evaluativa. Enseñar y aprender para la autonomía* (pp. 285-292). Barcelona: UAB-Editorial Síntesis.

Martín, E., García, L. A., Torbay, Á., & Rodríguez, T. (2008). Estrategias de aprendizaje y rendimiento académico en estudiantes universitarios. *International Journal of Psychology and Psychological Therapy*, 8(3), 401-412. Recuperado de <http://www.redalyc.org/articulo.oa?id=56080312>

Martínez, M. (2006). Reflexiones sobre la conveniencia de un cambio de cultura docente en la universidad. En M. Martínez & S. Carrasco (Coords.), *Propuestas para el cambio docente en la universidad* (pp. 15-32). Barcelona: Octaedro-ICE UB.

Martínez, L. F, Castejón, F. J., & Santos, M. L. (2012). Diferentes percepciones sobre la evaluación formativa entre profesorado y alumnado en la formación inicial de educación física. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 15(4), 57-67. Recuperado de http://www.aufop.com/aufop/uploaded_files/articulos/1364431164.pdf

Martínez, N., García, M. L., & Porto, M. (2012). Las prácticas de evaluación empleadas en Pedagogía en la Universidad de Murcia. *Educatio Siglo XXI*, 30(1), 275-296. Recuperado de <http://revistas.um.es/educatio/article/view/149241>

- Martínez, M .M., García, B. & Quintanal, J. (2006). El perfil del profesorado universitario de calidad desde la perspectiva del alumnado. *Educación XX1*, 9, 183-198. Recuperado de <http://revistas.uned.es/index.php/educacionXX1/article/view/325/281>
- Martínez, M., & Viader, M. (2008). Reflexiones sobre aprendizaje y docencia en el actual contexto universitario. La promoción de equipos docentes. *Revista de Educación*, nº extraordinario, 213-234. Recuperado de http://www.revistaeducacion.mec.es/re2008/re2008_09.pdf
- Martínez de la Hidalga, Z., & Villardón, L. (2014). La imagen profesional del profesorado en formación continua. *Revista Fuentes*, 15, 187-212. doi: 10.12795/revistafuentes.2014.i15.09
- Martínez-Mínguez, L., & Flores, G. (2014). Profesorado y egresados ante los sistemas de evaluación del alumnado en la formación inicial en la formación inicial del maestro en educación infantil. *Revista Digital de Investigación en Docencia Universitaria*, 1, 32-50. Recuperado de <http://revistas.upc.edu.pe/index.php/docencia/article/view/371/344>
- Martínez-Rizo, F., & Mercado, A. (2015). Estudios sobre prácticas de evaluación en el aula: Revisión de la literatura. *Revista Electrónica de Investigación Educativa*, 17(1), 17-32. Recuperado de <http://redie.uabc.mx/vol17no1/contenido-mtnzrizo-mercado.html>
- Marton, F., & Säljö, R. (2005). Approaches to learning. En F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning. Implications for teaching and studying in higher education* (3rd ed.) (pp. 39-58). Recuperado de http://www.docs.hss.ed.ac.uk/iad/Learning_teaching/Academic_teaching/Resources/Experience_of_learning/EoLChapter3.pdf (Trabajo original publicado en 1984).

- Mas, O. (2011). El profesor universitario: Sus competencias y formación. *Profesorado. Revista de currículum y formación del profesorado*, 15(3), 195-201. Recuperado de <http://www.ugr.es/~recfpro/rev153COL1.pdf>
- Mas, O. (2012). Las competencias del docente universitario: La percepción del alumno, de los expertos y del propio protagonista. *Revista de Docencia Universitaria*, 10(2), 299-318. Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/381>
- Mas, O., & Ruiz, C. (septiembre, 2007). *El profesor universitario en el nuevo espacio europeo de educación superior. Perfil competencial y necesidades formativas*. Trabajo presentado en el I Congreso Internacional de Nuevas tendencias en la formación permanente del profesorado. Barcelona. Recuperado de <http://tecnologiaedu.us.es/mec2011/htm/mas/3/31/48.pdf>
- Mateo, J. (2000). La evaluación del profesorado y la gestión de la calidad en educación. Hacia un modelo comprensivo de evaluación sistemática de la docencia. *Revista de Investigación Educativa*, 18(1), 7-34. Recuperado de <http://revistas.um.es/rie/article/view/121451/114141>
- Mateo, J. (2006a). Claves para el diseño de un nuevo marco conceptual para la medición y evaluación educativas. *Revista de Investigación Educativa*, 24(1), 165-186. Recuperado de <http://revistas.um.es/rie/article/view/97361>
- Mateo, J. (2006b). Nuevo enfoque de la evaluación de los aprendizajes en el contexto europeo de educación superior. En M. Martínez & S. Carrasco (Coords.), *Propuestas para el cambio docente en la universidad* (pp. 123-136). Barcelona: Octaedro-ICE UB.
- Mateo, J. (2008). Nuevo enfoque de evaluación de los aprendizajes en el contexto europeo de la educación superior. En J. Mateo, J. & F. Martínez (Eds.), *La evaluación*

alternativa de los aprendizajes (Cuadernos de docencia universitaria, 3) (pp. 5-22).

Recuperado de <http://www.octaedro.com/ice/pdf/DIG103.pdf>

Mateo, J. (Coord.). (2009a). *Guía para la evaluación de competencias en el Trabajo de Fin de Grado en el ámbito de las Ciencias Sociales y Jurídicas*. Recuperado del sitio de Internet de la Agència per a la Qualitat del Sistema Universitari de Catalunya (AQU): http://www.aqu.cat/doc/doc_30156918_1.pdf

Mateo, J. (2009b). La investigación ex post-facto. En R. Bisquerra (Coord.), *Metodología de la investigación educativa* (2ª ed.) (pp. 195-230). Madrid: La Muralla.

Mateo, J. (2012). La formación de formadores en la educación superior. *Revista de Docencia Universitaria*, 10(2), 211-223. Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/387>

Mateo, J., & Vlachopoulos, D. (2013). Evaluación en la universidad en el contexto de un nuevo paradigma para la educación superior. *Educación XX1*, 16(2), 183-208. doi: 10.5944/educxx1.16.2.2639

Mateos, M., & Pérez, M. P. (2006). El cambio de las concepciones de los alumnos sobre el aprendizaje. En J. I. Pozo et al. (Eds.), *Nuevas formas de pensar la enseñanza y el aprendizaje. Concepciones de profesores y alumnos* (pp. 403-417). Barcelona: Graó.

Mauri, T., Clarà, M., Ginesta, A., & Colomina, R. (2013). La contribución al aprendizaje en el lugar de trabajo de los equipos docentes universitarios. Un estudio exploratorio. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, 36(3), 341-360. doi: 10.1174/021037013807533025

McDonald, B. (2008). Evaluación y control de la educación. En J. Gimeno & A. I. Pérez (Eds.), *La enseñanza: Su teoría y su práctica* (6ª ed.) (pp. 467-478). Madrid: Akal.

- McDonald, B. (2010). Improving learning through meta assessment. *Active Learning in Higher Education*, 11(2), 119-129. doi: 10.1177/1469787410365651
- McDonald, R; Boud, D; Francis, J; & Gonczi, A. (2000). Nuevas perspectivas sobre la evaluación. *Boletín Cintefor. Boletín Técnico Interamericano de Formación Profesional*, 149, 41-72. Recuperado de <http://dialnet.unirioja.es/servlet/articulo?codigo=2122854>
- McDowell, L. (1995). Impact of innovative assessment. *Innovations in Education & Training International*, 32, 302-313. doi: 10.1080/1355800950320402
- McDowell, L., & Sambell, K. (2003). La experiencia en evaluación innovadora. En S. Brown & A. Glasner (Eds.), *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 91-103). Madrid: Narcea.
- McGarr, O., & Clifford, A. M. (2013). 'Just enough to make you take it seriously': Exploring students' attitudes towards peer assessment. *Higher Education*, 65, 677-693. doi: 10.1007/s10734-012-9570-z
- McMahon, T. (2010). Peer feedback in a undergraduate programme: Using action research to overcome students' reluctance to criticise. *Educational Action Research*, 18(2), 273-288. doi: 10.1080/09650791003741814
- McMillan, J. H. (2000). Fundamental assessment principles for teacher and school administrators. *Practical Assessment, Research & Evaluation*, 7(8). Recuperado de <http://pareonline.net/getvn.asp?v=7&n=8>
- Meade, D. (1997). El profesor de calidad. En P. Apodaca & C. Lobato (Eds.), *Calidad en la universidad: Orientación y evaluación* (pp. 206-222). Barcelona: Laertes.

- Medina, A., Domínguez, M. C., & Ribeiro, F. (2011). Formación del profesorado universitario en competencias docentes. *Revista Historia de la Educación Latinoamericana*, 13(17), 119-138. Recuperado de <http://www.redalyc.org/pdf/869/86922615006.pdf>
- Medina, M., & Verdejo, A. (2000). *La evaluación del aprendizaje estudiantil* (2ª ed.). San Juan: Isla Negra Editores.
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25). Recuperado de <http://pareonline.net/getvn.asp?v=7&n=25&cmd=View+Article>
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12(2), 101-112. doi: 10.1177/1365480209105575
- Meyer, J. H. (1992a). Study orchestration: The manifestation, interpretation, and consequences of contextualised approaches to learning. *Higher Education*, 22, 297-316.
- Meyer, J. H. (1992b). What's the difference between authentic and performance assessment. *Educational Leadership*, 49(8), 39-40.
- Meyer, J. H., & Watson, R. M. (1991). Evaluating the quality of student learning. II Study orchestration and the curriculum. *Studies in Higher Education*, 16(3), 251-275. doi: 10.1080/03075079112331382835
- Michavila, F. (2005). No sin los profesores. *Revista de Educación*, 337, 37-49. Recuperado de http://www.revistaeducacion.mec.es/re337_03.htm
- Michavila, F. (2011). Bolonia en crisis. *Revista de Docencia Universitaria*, 9(3), 15-28. Recuperado de http://red-u.net/redu/documentos/volumenes_completos_pdf/vol9_n3_completo.pdf

Miguel de, M. (2000). La evaluación de programas sociales. Fundamentos y enfoques teóricos. *Revista de Investigación Educativa*, 18(2), 289-318. Recuperado de <http://revistas.um.es/rie/article/view/121001/113691>

Miguel de, M. (2010). La evaluación como eje del cambio del EEES. En *Primeras jornadas internacionales sobre el EEES: Evaluación* (pp. 13-16). Barcelona: Universidad Internacional de Catalunya.

Miller, M. (1990). The assessment of clinical skills/competente/performance. *Academic Medicine*, 65(9), 563-567. Recuperado de <http://winbev.pbworks.com/f/Assessment.pdf>

Milligan, G. W., Wong, D. S., & Thompson, P. A. (1987). Robustness properties of nonorthogonal analysis of variance. *Psychological Bulletin*, 101, 464-470. doi: 10.1037/0033-2909.101.3.464

Ministerio de Educación. (2010). *Estrategia Universidad 2015: Contribución de las universidades al progreso socioeconómico español*. Recuperado de <http://www.mecd.gob.es/dctm/eu2015/2011-estrategia-2015-espanol.pdf?documentId=0901e72b80910099>

Ministerio de Educación. (2011). *Propuesta (Borrador). Estatuto del personal docente e investigador de las universidades públicas españolas. Versión informada por el Consejo de Universidades de 25 de mayo de 2011*. Recuperado de http://www.uv.es/ugt/lou/sexta_borrador_estatuto_pdi.pdf

Molenaar, W. M., Zantig, A., Van Beukelen, P., De Grave, W., Baane, J. A., Bustraán, J. A., ... Vervoorn, J. M. (2009). A framework of teaching competencies across the medical education continuum. *Medical Teacher*, 31, 390-396. doi: 10.1080/01421590902845881

- Molloy, E., & Boud, D. (2013). Changing conceptions of feedback? En D. Boud & E. Molloy (Eds.), *Effective feedback in higher and professional education. Understanding it and doing it well* (pp. 11-33). Abingdon: Routledge.
- Monedero, J. J. (1998). *Bases teóricas para la evaluación educativa*. Málaga: Aljibe.
- Monereo, C. (2003). La evaluación del conocimiento estratégico a través de tareas auténticas. *Revista Pensamiento Educativo. Revista de Investigación Educativa Latinoamericana*, 32, 71-89. Recuperado de <http://pensamientoeducativo.uc.cl/index.php/pel/article/view/246/516>
- Monereo, C. (2009). La autenticidad de la evaluación. En M. Castelló (Coord.), *La evaluación auténtica en enseñanza secundaria y universitaria: Investigación e innovación* (pp. 15-32). Barcelona: Edebé.
- Monereo, C. (2013). La investigación en la formación del profesorado universitario: Hacia una perspectiva integradora. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, 36(3), 281-291. doi: 10.1174/021037013807533052
- Monereo, C. (2014a). Las competencias del profesor universitario del siglo XXI: Hacia una identidad profesional docente. En C. Monereo (Coord.), *Enseñando a enseñar en la universidad. Formación del profesorado basada en incidentes críticos* (pp. 39-61). Barcelona: Octaedro.
- Monereo, C. (2014b). Salir de la crisis. Sistemas de formación basados en incidentes críticos. En C. Monereo (Coord.), *Enseñando a enseñar en la universidad. Formación del profesorado basada en incidentes críticos* (pp. 93-127). Barcelona: Octaedro.

- Monereo, C., & Badía, A. (2011). Los heterónimos del docente: Identidad, selfs y enseñanza. En C. Monereo & J. I. Pozo (Eds.), *La identidad en Psicología de la Educación: Necesidad, utilidad y límites* (pp. 57-76.). Madrid: Narcea.
- Monereo, C., & Castelló, M. (2009a). Las competencias del profesor para evaluar y su evaluación. En C. Monereo (Coord.), *Pisa como excusa. Repensar la evaluación para cambiar la enseñanza* (pp. 73-90). Barcelona. Graó.
- Monereo, C., & Castelló, M (2009b). La evaluación como herramienta de cambio educativo: Evaluar las evaluaciones. En C. Monereo (Coord.), *Pisa como excusa. Repensar la evaluación para cambiar la enseñanza* (pp. 13-30). Barcelona: Graó.
- Monereo, C., & Domínguez, C. (2014). La identidad docente de los profesores universitarios competentes. *Educacion XX1*, 17(2), 83-104. doi: 10.5944/educxx1.17.2.11480
- Monereo, C., & Lemus, R. (2010). Dime cómo evalúas y te diré cómo aprenden (a aprender) tus alumnos. *Aula de Innovación Educativa*, 190, 56-59.
- Monereo, C., Weise, C., & Alvarez, I. (2013). Cambiar la identidad docente en la universidad. Formación basada en incidentes dramatizados. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, 36(3), 323-340. doi: 10.1174/021037013807533043
- Monescillo, M., & Toscano, M. O. (2007). Mejorar el proceso evaluador en el contexto universitario. En M. C. Fonseca & J. I. Aguaded (Dir.), *Enseñar en la universidad. Experiencias y propuestas para la docencia universitaria* (pp. 51-67). La Coruña: Netbiblo.
- Monroy, F., & Hernández, F. (2014). Factores que influyen en los enfoques de aprendizaje universitario. Una revisión sistemática. *Educación XX1*, 17(2), 105-124. doi: 10.5944/educxx1.17.2.11481

- Montero, M. L. (2006). Profesores y profesoras en un mundo cambiante: El papel clave de la formación inicial. *Revista de Educación*, 340, 66-86. Recuperado de http://www.revistaeducacion.mec.es/re340_02.html
- Moore, S., & Murphy, M. (2009). *Estudiantes excelentes. 100 ideas prácticas para mejorar el autoaprendizaje en educación superior*. Madrid. Narcea.
- Mora, J. G. (2001). The academic profession in Spain: Between the civil service and the market. *Higher Education*, 41, 131-155.
- Morales, P. (2006a). Implicaciones para el profesor de una enseñanza centrada en el alumno. *Miscelánea Comillas*, 64(124), 11-28. Recuperado de <http://web.upcomillas.es/personal/peter/otrosdocumentos/Implicaciones%20profesor%20ense%F1anza%20centrada%20alumno.pdf>
- Morales, P. (2006b). *Medición de actitudes en psicología y en educación: Construcción de escalas y problemas metodológicos*. Madrid: Universidad Pontificia de Comillas.
- Morales, P. (2007). *La fiabilidad de los test y escalas* [Apuntes]. Recuperado de <http://www.upcomillas.es/personal/peter/estadisticabasica/Fiabilidad.pdf>
- Morales, P. (2008a). *Correlación y covarianza* [Apuntes]. Recuperado de <http://www.upcomillas.es/personal/peter/estadisticabasica/correlacion.pdf>
- Morales, P. (2008b). Estrategias para evaluar y calificar el producto del equipo: Cómo diferenciar las calificaciones individuales. En L. Prieto (Coord.), *La enseñanza universitaria centrada en el aprendizaje. Estrategias útiles para el profesorado* (pp. 151-169). Barcelona: Octaedro-ICE UB.
- Morales, P. (2008c). Nuevos roles de profesores y alumnos, nuevas formas de enseñar y de aprender. En L. Prieto (Coord.), *La enseñanza universitaria centrada en el*

aprendizaje. Estrategias útiles para el profesorado (pp. 17-29). Barcelona: Octaedro-ICE UB.

Morales, P. (2012). *El tamaño del efecto (effect size): Análisis complementario al contraste de medias* [Apuntes]. Recuperado de <http://web.upcomillas.es/personal/peter/investigacion/Tama%F1oDelEfecto.pdf>

Morales, P. (2013). *El análisis factorial en la construcción e interpretación de test, escalas y cuestionarios* [Apuntes]. Recuperado de <http://www.upcomillas.es/personal/peter/investigacion/AnalisisFactorial.pdf>

Morales-Olivas, F. J., & Estañ, L. (2007). Evaluación de una asignatura optativa de la licenciatura de periodismo. ¿Se puede evaluar sin examinar? En B. Salinas & C. Cotillas (Coords.), *La evaluación de los estudiantes en la educación superior. Apuntes de buenas prácticas* (pp. 37-39). Recuperado de <http://www3.uji.es/~betoret/Formacion/Evaluacion/Documentacion/La%20evaluacion%20estudiantes%20en%20la%20ESuperior%20UV.pdf>

Moril, R., Ballester, L., & Martínez, J. (2012). Introducción de las valoraciones de valoración analítica en el proceso de evaluación del Practicum de los Grados de Infantil y Primaria. *Revista de Docencia Universitaria*, 10(2), 351-372. Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/229>

Morse, D. T. (1998). Minsize: A computer program for obtaining minimum sample size as an indicator of effect size. *Educational and Psychological Measurement*, 58, 150-161. doi: 10.1177/0013164498058001012

Morse, D. T. (1999). Minsize2: A computer program for determining effect size and minimum sample size for statistical significance for univariate, multivariate, and nonparametric tests. *Educational and Psychological Measurement*, 59, 518-531. doi: 10.1177/00131649921969901

- Mueller, J. (2004). *Authentic Assessment Toolbox*. Recuperado de <http://jfmueLLer.faculty.noctrl.edu/toolbox/whatisit.htm#definitions>
- Mulder, M., Gulikers, J., Biemans, H., & Wesselink, R. (2009). The new competence concept in higher education: Error or enrichment? *Journal of European Industrial Training*, 33(8/9), 755-770. doi: 10.1108/03090590910993616
- Murillo, F. J. (2008). La evaluación del profesorado universitario en España. *Revista Iberoamericana de Evaluación Educativa*, 1(3), 29-45. Recuperado de http://rinace.net/riee/numeros/vol1-num3_e/art3.pdf
- Murphy, R. (2006). Evaluating new priorities for assessment in higher education. En C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 37-47). New York: Routledge.
- Muñoz, G. (2007). Un nuevo paradigma: 'La quinta generación de evaluación'. *Laurus*, 15(23), 158-198. Recuperado de <http://www.redalyc.org/articulo.oa?id=76102309>

N

- Nacional Center for Research on Evaluation, Standards & Student Testing (CRESST). (2012). *Stepping up 2012: Advancing assessment, evaluation and technology Worldwide*. Recuperado de http://www.cse.ucla.edu/about/Overview_2012.pdf
- National Union Students (NUS). (2010). *Charter on feedback and assessment*. Recuperado de: <http://www.nusconnect.org.uk/news/article/highereducation/720/>
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22(2), 155. doi: 10.1207/s15326985ep2202_4

Navaridas, F. (2002). La evaluación del aprendizaje y su influencia en el comportamiento estratégico del estudiante universitario. *Contextos Educativos*, 5, 141-156.

Recuperado de

<https://publicaciones.unirioja.es/ojs/index.php/contextos/article/view/509>

Navarro, V., Santos, M. L., Buscà, F., Martínez, L., & Martínez, L. F. (2010). La experiencia de la Red Universitaria Española de Evaluación Formativa y Compartida: Proceso y abordaje. *Revista Iberoamericana de Educación*, 52(7). Recuperado de

<http://www.rieoei.org/expe/3428Navarro.pdf>

Navío, A. (2005). *Las competencias profesionales del formador. Una visión desde la formación continua*. Barcelona: Octaedro.

Nevada State College (NSC), Outcomes Assessment Committee. (2009). *Glossary of assessment and evaluation terms*. Recuperado de <http://nsc.nevada.edu/1289.asp>

Nevo, D. (1997). *Evaluación basada en el centro. Un diálogo para la mejora educativa*.

Bilbao: Ediciones Mensajero.

Nicol, D. (mayo, 2007). *Principles of good assessment and feedback: Theory and practice*.

Paper presentado en la REAP International online Conference on Assessment Design for Learner Responsibility. Recuperado de

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.366.4201&rep=rep1&type=pdf>

Nicol, D. (2009). *Transforming assessment and feedback: Enhancing integration and empowerment in the first year*. Recuperado de

http://dera.ioe.ac.uk/11605/1/First_Year_Transforming_Assess.pdf

- Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35(5), 501–517. doi: 10.1080/02602931003786559
- Nicol, D (2013). Resituating feedback from the reactive to the proactive. En D. Boud & L. Molloy (Eds.), *Effective feedback in higher and professional education. Understanding it and doing it well* (pp. 34-49). Abingdon: Routledge.
- Nicol, D., Breslin, C., & Thomson, A. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102-122. doi: 10.1080/02602938.2013.795518
- Nicol, D., & Macfarlane, D. (2004). *Rethinking formative assessment in higher education: A theoretical model and seven principles of good feedback practice*. Recuperado de http://www-new1.heacademy.ac.uk/assets/documents/assessment/web0015_rethinking_formative_assessment_in_he.pdf
- Nicol, D., & Macfarlane, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218. doi: 10.1080/03075070600572090
- Nieto, S. (2000). El discurso del profesorado universitario sobre la evaluación del aprendizaje de los alumnos como estrategia de innovación y cambio profesional: Exposición y análisis de una experiencia. *Revista de Educación*, 322, 305-324.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Núñez, J. C., González, J. A., Rodríguez, M. G., González-Pumariiega, S., Montero, C. R., Pérez, L. A., & Torres, M. D. (1998). Estrategias de aprendizaje, autoconcepto y

rendimiento académico. *Psicoterma*, 10(1), 97-109. Recuperado de <http://www.psicothema.com/pdf/146.pdf>

Núñez, J. L., & Reyes, C. I. (2014). La evaluación del aprendizaje de estudiantes: Validación española del Assessment Experience Questionnaire (AEQ). *Estudios Sobre Educación*, 26, 63-77. Recuperado de <http://dadun.unav.edu/handle/10171/36785>

O

Olivia, P., & Henson, K. (1980). What are the essential generic teaching competencies? *Theory into Practice* 9 (2), 117-121. doi: 10.1080/00405848009542884

Olmos, S. (2008). *Evaluación formativa y sumativa de estudiantes universitarios: Aplicación de las tecnologías a la evaluación educativa* (Tesis doctoral, Universidad de Salamanca). Recuperado de <http://hdl.handle.net/11162/89282>

Olmos, S., & Rodríguez, M. J. (2010). Diseño del proceso de evaluación de los estudiantes universitarios españoles: ¿Responde a una evaluación por competencias en el Espacio Europeo de Educación Superior? *Revista Iberoamericana de Educación*, 53(1). Recuperado de <http://www.rieoei.org/deloslectores/3757Olmos.pdf>

Orden ECI/3854/2007, de 27 de diciembre, por la que se establecen los requisitos para la verificación de los títulos universitarios oficiales que habiliten para el ejercicio de la profesión de Maestro en Educación Infantil.

Orden ECI/3857/2007, de 27 de diciembre, por la que se establecen los requisitos para la verificación de los títulos universitarios oficiales que habiliten para el ejercicio de la profesión de Maestro en Educación Primaria.

- Orden de la, A. (1985a). Modelos de evaluación universitaria. *Revista española de pedagogía*, 43, 521-537.
- Orden de la, A. (1985b). Hacia una conceptualización del producto educativo. *Revista de Investigación Educativa*, 3(2), 271-283. Recuperado de <http://revistas.um.es/rie/article/view/136681/124261>
- Orden de la, A. (2011a). El problema de las competencias en la educación general. *Bordón*, 63(1), 47-63.
- Orden de la, A. (2011b). Reflexiones en torno a las competencias como objeto de evaluación en el ámbito educativo. *Revista Electrónica de Investigación Educativa*, 13(2). Recuperado de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1607-40412011000200001
- Orsmond, P., Maw, S. J., Park, J. R., Gomez, S., & Crook, A. (2013). Moving feedback forward: Theory to practice. *Assessment & Evaluation in Higher Education*, 38(2), 240-252. doi: 10.1080/02602938.2011.625472
- Orsmond, P., & Merry, S. (2011). Feedback alignment: Effective and ineffective links between tutors' and students' understanding of coursework feedback. *Assessment & Evaluation in Higher Education*, 36(2), 125-36. doi: 10.1080/02602930903201651
- Ortega, M .C. (2010). Competencias emergentes del docente ante las demandas del Espacio Europeo de Educación Superior. *Revista Española de Educación Comparada*, 16, 305-327. Recuperado de http://www.uned.es/reec/pdfs/16-2010/14_ortega.pdf

Ortiz, E. (2006). Modelo de evaluación del aprendizaje. *Paideia*, 1(2). Recuperado de <http://paideia.uprrp.edu/wp-content/uploads/2013/11/Modelo-de-evaluaci%C3%B3n-del-aprendizaje.pdf>

Ory de, M., & Ruiz, V. (2011). La evaluación en el aula de primaria. Factor clave para el aprendizaje de las ciencias y las matemáticas. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 8(2), 212-220. doi: 10498/10856

Osney Grange Grupo (OGG), ASKe (Assessment Standards Knowledge exchange). (2009). *Feedback: An agenda for change*. Recuperado de <http://www.brookes.ac.uk/aske/documents/ManifestoLeafletNew.pdf>

Otero, L. (2000). Saber/saber hacer: Una propuesta para mejorar la evaluación. En M. Ballester et al. (Eds.), *Evaluación como ayuda al aprendizaje* (pp. 151-154). Barcelona: Graó.

P

Padilla, M. T., Gil, J., Rodríguez, J., Torres, J. J., & Clarés, J. (2010). Evaluando el sistema de evaluación del aprendizaje universitario: Análisis documental de la Universidad de Sevilla. *Revista Iberoamericana de Educación*, 53(3). Recuperado de <http://www.rieoei.org/deloslectores/3574Padilla.pdf>

Pajares, M. E. (1992). Teachers beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307-332. doi: 10.3102/00346543062003307

- Palacios, A., & V. M. López (2013). Haz lo que digo pero no lo que hago: Sistemas de evaluación del alumnado en la formación inicial del profesorado. *Revista de Educación*, 361, 279-305. doi: 10.4438/1988-592X-RE-2011-361-143
- Palacios, A., López, V. M., & Barba, J. J. (2013). Tipologías de profesorado universitario en función de la evaluación aplicada a los futuros docentes. *Estudios Sobre Educación*, 24, 173-195. Recuperado de <http://dadun.unav.edu/handle/10171/29569>
- Palomares, A. (2011). El modelo docente universitario y el uso de nuevas metodologías en la enseñanza, aprendizaje y evaluación. *Revista de Educación*, 355, 591-604. doi: 10-4438/1988-592X-RE-2011-355-038
- Panadero, E., & Jonsson, A. (2013). The use of scouring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 123-144. doi: 10.1016/j.edurev.2013.01.002
- Parcerisa, A. (2006). Plan docente: Planificar las asignaturas en el marco europeo de educación superior. En M. Martínez & S. Carrasco (Coords.), *Propuestas para el cambio docente en la universidad* (pp. 49-74). Barcelona: Octaedro-ICE UB
- Parcerisa, A. (Coord.). (2010). *Ejes para la mejora docente en la universidad*. Barcelona: Octaedro-ICE UB.
- Paris, D. (2013). The last artisans? Traditional and future faculty roles. *Peer Review*, summer 2013, 17-20. Recuperado de <http://www.aacu.org/peerreview/2013/summer/paris>
- Patel, M. (2002). A meta-evaluation, or quality assessment, of the evaluation in this issue, based on the African Evaluation Guidelines: 2002. *Evaluation and Program Planning*, 25, 293-332. doi: 10.1016/S0149-7189(02)00043-5
- Patton, M. Q. (1982). *Practical evaluation*. London: Sage.

- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Beverly Hills: Sage Publications.
- Pavié, A. (2011). Formación docente: Hacia una definición del concepto de competencia profesional docente. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 14(1), 67-80. Recuperado de http://www.aufop.com/aufop/uploaded_files/articulos/1301587967.pdf
- Pearson, D., Vyas, S., Sensale, L., & Kim, Y. (2001). Making our way through the assessment and accountability maze: Where do we go now? *The Clearing House*, 74(4), 175–91. doi: 10.1080/00098650109599186
- Perales, M. J., Sánchez, P., & Chiva, I. (2002). El `curso de iniciación a la docencia universitaria´ como experiencia de formación de profesores universitarios noveles en la Universitat de València. Un sistema de evaluación. *Revista Electrónica de Investigación y Evaluación Educativa*, 8(1), 49-69. Recuperado de http://www.uv.es/RELIEVE/v8n1/RELIEVEv8n1_4.pdf
- Pérez, A. I. (2010a). Aprender a educar. Nuevos desafíos para la formación de docentes. *Revista Interuniversitaria de Formación del Profesorado*, 68(24,4), 37-60.
- Pérez, A. I. (2010b). Nuevas exigencias y escenarios para la profesión docente en la era de la información y la incertidumbre. *Revista Interuniversitaria de Formación del Profesorado*, 68(24,4), 17-36.
- Pérez, G. (1986). Crítica al concepto de rendimiento académico. *Revista española de pedagogía*, 174, 521-534.
- Pérez, G. (1994). *Investigación cualitativa. Retos e interrogantes. II. Técnicas y análisis de datos*. Madrid: La Muralla.

Pérez Juste¹⁰³, R. (Dir.). (1991). *La universidad Nacional de Educación a Distancia*.

Aproximación a la evaluación de un modelo innovador. Madrid: CIDE.

Pérez Juste, R. (Dir.). (2002). La evaluación de programas en el marco de la evaluación de calidad. *XXI, Revista de Educación*, 4, 43-76.

Pérez Pérez¹⁰⁴, R. (2004). Mida las tareas de aprendizaje. En L. M. Villar (Coord.), *Programa para la mejora de la docencia universitaria* (pp. 523-541). Madrid: Pearson-Prentice Hall.

Pérez, M. L., Juandó, J., & Carretero, R. (2009). Formación de profesores y evaluación. IRAESO: Un instrumento para reflexionar sobre la evaluación del aprendizaje. En M. Castelló (Coord.), *La evaluación auténtica en enseñanza secundaria y universitaria: Investigación e innovación* (pp. 269-292). Barcelona: Edebé.

Pérez, A., Julián, J. A., & López, V. M. (2009). Evaluación formativa y compartida en el EEES. En V. M. López (Coord.), *Evaluación formativa y compartida en educación superior. Propuestas, técnicas, instrumentos y experiencias* (pp. 19-44). Madrid: Narcea.

Pérez, M. L., Reyes, M., Palma, M., & Rafel, E. (2000). La evaluación de la calidad del aprendizaje en la universidad. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, 23(91), 5-30. doi: 10.1174/021037000760087847

Pérez, F., Serrano, L., Pastor, J. M., Hernández, L., Soler, A., & Zaera, I. (Dirs.). (2012). *Universidad, universitarios y productividad en España*. Recuperado del sitio de

¹⁰³ Excepcionalmente la referencia incluye ambos apellidos por ser la única alternativa de para diferenciarlo de Pérez Pérez, ambos llamados Ramón.

¹⁰⁴ Excepcionalmente la referencia incluye ambos apellidos por ser la única alternativa de para diferenciarlo de Pérez Juste, ambos llamados Ramón. La única forma de diferenciar con Pérez Juste (también se llama Ramón).

Internet de la Fundación BBVA:

http://www.fbbva.es/TLFU/dat/Monografia_Universidades_2012_web_cerrado.pdf

Pérez, A., Tabernero, B., López, V. M., Ureña, N., Ruiz, E., Capllonch, M., ... Castejón, J. (2008). Evaluación formativa y compartida en la docencia universitaria: La concreción de cuestiones/clave para su aplicación en el camino hacia el Espacio Europeo de Educación Superior (EEES). *Revista de Educación*, 347, 435-451.

Recuperado de http://www.revistaeducacion.mec.es/re347_20.html

Perrenoud, P. (2004). *Diez nuevas competencias para enseñar. Invitación al viaje*. Barcelona: Graó.

Perrenoud, P. (2010). La formación del profesorado: Un compromiso entre visiones inconciliables de la docencia. *Revista Interuniversitaria de Formación del Profesorado*, 68(24,4), 103-122.

Poblete, M. (2007). Evaluación de competencias en la educación superior. En *III Jornadas Universitarias de Innovación y Calidad: Buenas Prácticas Académicas para la Innovación del Proceso de Aprendizaje en el Espacio Europeo de Educación Superior (EEES)* [CD]. Bilbao: Universidad de Deusto.

Ponte, J. P. (1994). Mathematics teachers' professional knowledge. En J. Ponte & J. Mateos (Eds.), *Proceedings of the Eighteenth International Conference for the Psychology of Mathematics Education* (Vol. I) (pp. 195-210). Lisboa: International Group for the Psychology of Mathematics Education. Recuperado de la base de datos ERIC. (ED383537)

Pope, N. (2005). The impact of stress in self-and peer assessment. *Assessment & Evaluation in Higher Education*, 30(3), 51–63. doi: 10.1080/0260293042000243896

Popham, W. J. (1980). *Problemas y técnicas de la evaluación educativa*. Madrid: Anaya.

- Porto, M. (2005a). Evaluación de estudiantes y calidad docente en la Universidad de Santiago de Compostela: Propuestas de mejora. *Revista de Enseñanza Universitaria*, 26, 7-17. Recuperado de <http://institucional.us.es/revistas/universitaria/26/01%20porto.pdf>
- Porto, M. (2005b). *La evaluación de los estudiantes universitarios. El caso de la Universidad de Santiago de Compostela* (Tesis doctoral). Recuperado de <https://dspace.usc.es/handle/10347/9684>
- Porto, M. (2006): La evaluación de estudiantes universitarios vista por sus protagonistas. *Educatio Siglo XXI*, 24, 167-187. Recuperado de <http://revistas.um.es/educatio/article/view/156/139>
- Porto M. (2008): Evaluación de estudiantes universitarios: ¿Toma de decisiones colegiada? *Cuadernos de Educación y Filosofía*, 34, 133-149. Recuperado de <http://www.scielo.org.ar/pdf/cfhycs/n34/n34a10>
- Porto, M. (2010). Inclusión de la evaluación de estudiantes como indicador de calidad institucional: Avances en los últimos diez años. *Revista Iberoamericana de Educación*, 52(3), 1-12. Recuperado de <http://www.rieoei.org/deloslectores/2911Porto.pdf>
- Porto, M., & Rodríguez, R. (2009). Evaluación de estudiantes en la Universidad de Santiago de Compostela: Percepciones de los propios alumnos. *Revista de Docencia Universitaria*, 9(2). Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/99>
- Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. *Assessment & Evaluation in Higher Education*, 33(2), 143-154. doi: 10.1080/02602930601127869

- Pozo, J. I. (2001). *Humana mente: El mundo, la conciencia y la carne*. Madrid: Morata.
- Pozo, J. I., Scheuer, N., Mateos, & Pérez, M. P. (2006). Las teorías implícitas sobre el aprendizaje y la enseñanza. En J. I. Pozo et al. (Eds.), *Nuevas formas de pensar la enseñanza y el aprendizaje. Concepciones de profesores y alumnos* (pp. 95-133). Barcelona: Graó.
- Prat, R., & Doval, E. (2005). Construcción y análisis de escalas. En J. P. Lèvy & J. Varela (Dirs.), *Análisis multivariable para las ciencias sociales* (pp. 44-89). Madrid: Pearson Prentice.
- Pratt, D. D. (1992). Conceptions of teaching. *Adult Education Quarterly*, 42, 203-220. doi: 10.1177/074171369204200401
- Prebble, T., Hargraves, H., Leach, L., Naidoo, K., Suddaby, G., & Zepke, N. (2004). *Impact of student support services and academic development programmes on student outcomes in undergraduate tertiary study: A synthesis of the research*. Recuperado del sitio de Internet del Ministerio de Educación de Nueva Zelanda: http://www.educationcounts.govt.nz/data/assets/pdf_file/0013/7321/ugradstudentoutcomes.pdf
- Price, M., Handley, K., Miller, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment & Evaluation in Higher Education*, 35(3), 277-289. doi: 10.1080/02602930903541007
- Price, M., & O'Donovan, B. (2006). Improving performance through enhancing student understanding of criteria and feedback. En C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 100-109). New York: Routledge.
- Prieto, L. (2004). La alineación constructivista en el aprendizaje universitario. En J. C. Torre & E. Gil (Eds.), *Hacia una enseñanza universitaria centrada en el aprendizaje. Libro*

homenaje a Pedro Morales (pp. 110-143). Madrid: Universidad Pontificia de Comillas.

Prieto, L. (2007). *Autoeficacia del profesor universitario. Eficacia percibida y práctica docente*. Madrid: Narcea.

Prieto, M., & Contreras, G. (2008). Las concepciones que orientan las prácticas evaluativas: Un problema a desvelar. *Estudios Pedagógicos*, 34(2), 245-262. Recuperado de <http://www.redalyc.org/articulo.oa?id=173514136014>

Prieto, G., & Delgado, A. R. (2010). Fiabilidad y validez. *Papeles del Psicólogo*, 31(1), 67-74. Recuperado de <http://www.papelesdelpsicologo.es/pdf/1797.pdf>

Prosser, M., & Trigwell, K. (1999). *Understanding learning and teaching. The experience in higher education*. Buckingham: Open University Press.

Q

Quality Assurance Agency for Higher Education (QAA). (2011). *UK Quality Code for higher education* (Part B: Assuring and enhancing academic quality. Chapter B6: Assessment of students and accreditation of prior learning) (3th ed.). Recuperado en agosto de 2011 de <http://www.qaa.ac.uk/Publications/InformationAndGuidance/Documents/Quality%20Code%20-%20Chapter%20B6.pdf>

Quality Assurance Agency for Higher Education (QAA). (2014). *UK Quality Code for higher education* (Part B: Assuring and enhancing academic quality. Chapter B6: Assessment of students and accreditation of prior learning) (4th ed.). Recuperado de <http://www.qaa.ac.uk/en/Publications/Documents/quality-code-B6.pdf>

Quality Assurance Agency for Higher Education (QAA). (2012a). *Outcomes from IQER: 2010-2011. Assessment*. Recuperado de <http://www.qaa.ac.uk/en/Publications/Documents/Outcomes-from-IQER-Assessment.pdf>

Quality Assurance Agency for Higher Education (QAA). (2012b). *Understanding assessment: Its role in safeguarding academic standards and quality in higher education. A guide for early career staff* (2nd ed.). Recuperado de <http://www.qaa.ac.uk/en/Publications/Documents/understanding-assessment.pdf>

Quesada, V., Rodríguez, G., & Ibarra, M. S. (2013). ActEval: Un instrumento para el análisis y reflexión de la actividad evaluadora del profesorado universitario. *Revista de Educación*, 362, 69-104. doi: 10-4438/1988-592X-RE-2011-362-153

R

Race, P. (2003). ¿Por qué evaluar de un modo innovador? En S. Brown & A. Glasner (Eds.), *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 77-90). Madrid: Narcea.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Sciences*, 28, 4-13. doi: 10.1002/bs.3830280103

Ramos, A .M. (2008). El Espacio Europeo de Educación Superior (EEES) y el Sistema Europeo de Transferencia de Créditos (ECTS). En A. Rodríguez, M. J. Caurcel, & A .M. Ramos (Coords.), *Didáctica en el Espacio Europeo de Educación Superior* (pp. 11-44). Madrid: Editorial EOS.

- Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education*, 16(2), 129-150. doi: 10.1080/03075079112331382944
- Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). London: Routledge Falmer.
- Real Academia Española (RAE). (2014). *Diccionario de la lengua española* (23ª ed.). Madrid: Espasa.
- Real Decreto 1044, de 1 de agosto de 2003, por el que se establece el procedimiento para la expedición por las universidades del Suplemento Europeo al Título.
- Real Decreto 1393/2007, de 29 de octubre, por el que se establece la ordenación de las enseñanzas universitarias oficiales.
- Real Decreto 1791/2010, de 30 de diciembre, por el que se aprueba el Estatuto del Estudiante Universitario.
- Real Decreto 1837/2008, de 8 de noviembre, por el que se incorporan al ordenamiento jurídico español la Directiva 2005/36/CE, del Parlamento Europeo y del Consejo, de 7 de septiembre de 2005, y la Directiva 2006/100/CE, del Consejo, de 20 de noviembre de 2006, relativas al reconocimiento de cualificaciones profesionales, así como a determinados aspectos del ejercicio de la profesión de abogado.
- Reber, A .S. (1993). *Implicit learning and tacit knowledge*. New York: Oxford University Press.
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448. doi: 10.1080/02602930902862859

Resolución de 4 de julio de 2012, de la Dirección General de Empleo, por la que se registra y publica el XIII Convenio colectivo de ámbito estatal para los centros de educación universitaria e investigación.

Riegle, R. P. (1987). Conceptions of faculty development. *Educational Theory*, 37, 53-59.
doi: 10.1111/j.1741-5446.1987.00053.x

Rhodes, G., & Tallantyre, F. (2003). Evaluación de las habilidades básicas. En S. Brown & A. Glasner (Eds.), *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 129-143). Madrid: Narcea.

Richardson, J. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education*, 30 (4), 387-415. doi: 10.1080/02602930500099193

Ricoy, M. C., & Fernández-Rodríguez, J. (2013). La percepción que tienen los estudiantes universitarios sobre la evaluación: Un estudio de caso. *Educación XX1*, 16(2), 321-342. doi: 10.5944/educxx1.16.2.2645

Rizo, H. E. (1999). Evaluación del docente universitario. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 2(1), 425-440. Recuperado de <http://aufop.com/aufop/revistas/arta/digital/100/653>

Rizo, H. E. (2004). La evaluación del aprendizaje: Una propuesta de evaluación basada en productos académicos. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 2(2), 19-29. Recuperado de <http://www.ice.deusto.es/RINACE/reice/vol2n2/Rizo.pdf>

Roach, P. (2003). Caso práctico: Primera evaluación por los compañeros y autoevaluación. En S. Brown & A. Glasner (Eds.), *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 211-220). Madrid: Narcea.

- Roberston, J., & Bond, C. (2008). Formas de ser en la universidad. En R. Barnett (Ed.), *Para una transformación de la universidad. Nuevas relaciones entre investigación, saber y docencia* (pp. 109-123). Barcelona: Octaedro.
- Rodrigo, M. J. (1993). Representaciones y procesos en las teorías implícitas. En M. J. Rodrigo, A. Rodríguez, & J. Marrero (Coords.), *Las teorías implícitas. Una aproximación al conocimiento cotidiano* (pp. 95-122). Madrid: Visor.
- Rodrigo, M. J. (1997). Del escenario sociocultural al constructivismo episódico: Un viaje al conocimiento escolar de la mano de las teorías implícitas. En M. J. Rodrigo & J. Arnay (Comp.), *La construcción del conocimiento escolar* (pp. 177-194). Barcelona: Paidós.
- Rodríguez, C. L. (2008). *Metaevaluación del proyecto evaluar para transformar* (Tesis doctoral). Universidad de Deusto, Bilbao.
- Rodríguez, G. (Coord.). (1998). La evaluación desde la perspectiva del alumnado universitario: Propuestas para la mejora. *Revista de Enseñanza Universitaria*, número extraordinario, 223-234.
- Rodríguez, J. (2008). Las competencias de evaluación de la clase magisterial del futuro. *Investigación y Evaluación Educativa*, 4(2), 7-10. Recuperado de <http://ineva.uprrp.edu/boletin/v0004n0002.pdf>
- Rodríguez, S. (1982). *Factores de rendimiento escolar*. Barcelona: Oikos-Tau.
- Rodríguez, S. (2003). Nuevos retos y enfoques en la formación del profesorado universitario. *Revista de Educación*, 331, 67-99.
- Rodríguez, S. (2005). La gestión de la calidad en el marco de la enseñanza universitaria. En W. De Vries (Coord.), *Calidad eficiencia y evaluación de la educación superior* (pp. 243-263). La Coruña: Netbiblo.

- Rodríguez, G., & Ibarra, M. S. (Eds.). (2011). e-Evaluación orientada al e-Aprendizaje estratégico en la educación superior. Madrid: Narcea.
- Rodríguez, G., & Ibarra, M. S. (2012). Reflexiones en torno a la competencia evaluadora del profesorado en la Educación Superior. *Revista de Docencia Universitaria* [Monográfico: Competencias docentes en la educación superior], 10(2), 149-161. Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/394>
- Rodríguez, G., Ibarra, M., Gallego, B., Gómez, M. A., & Quesada, V. (2012). La voz del estudiante en la evaluación del aprendizaje: Un camino por recorrer en la universidad. *Revista Electrónica de Investigación y Evaluación Educativa*, 18(2). doi: 10.7203/relieve.18.2.1985
- Román, M., & Murillo, F. J. (2009). La evaluación de los aprendizajes: Un recurso estratégico para mejorar la calidad educativa. *Revista Iberoamericana de Evaluación Educativa*, 2(1), 5-9. Recuperado de <http://www.rinace.net/riee/numeros/vol2-num1/editorial.html>
- Romero, M. P. (2004). Evalúe formativa y sumativamente. En L. M. Villar (Coord.), *Programa para la mejora de la docencia universitaria* (pp. 509-522). Madrid: Pearson-Prentice Hall.
- Romero, R., Fraile, A., López, V. M., & Castejón, F. J. (2014). The relationship between formative assessment systems, academic performance and teacher and student workloads in higher education. *Infancia y Aprendizaje. Journal for the Study of Education and Development*, 37(2), 310-341. doi: 10.1080/02103702.2014.918818
- Rotger, B. (1992). *Evaluación formativa* (Serie Educación y futuro, vol. 9). Madrid: Cincel.

- Rubio, M. J., & Berlanga, V. (2012). Cómo aplicar las pruebas paramétricas bivariadas t de student y ANOVA en SPSS. Caso práctico. *Revista d'Innovació i Recerca en Educació*, 5(2), 83-100. doi:10.1344/reire2012.5.2527
- Rué, J. (2007). *Enseñar en la universidad. El EEES como reto para la educación superior*. Madrid: Narcea.
- Rué, J. (2008). Formar en competencias en la universidad: Entre la relevancia y la banalidad. *Revista de Docencia Universitaria*, 6(1). Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/58/pdf>
- Rué, J. (2009a). El cambio en la universidad, sus epistemologías y consecuencias de las mismas. *Revista Complutense de Educación*, 20(2), 295-317. Recuperado de <http://revistas.ucm.es/index.php/RCED/article/view/RCED0909220295A/15288>
- Rué, J. (Coord.). (2009b). *Un model d'anàlisi de la qualitat de l'ensenyament des de la perspectiva dels estudiants a titulacions de quatre universitats catalanes (UAB, UB, UPC, URV)*. Recuperado del sitio de Internet de la Agència per a la Qualitat del Sistema Universitari de Catalunya (AQU): http://www.aqu.cat/doc/doc_15053340_1.pdf
- Rueda, B. (2009). La evaluación del desempeño docente: Consideraciones desde el enfoque por competencias. *Revista Electrónica de Investigación Educativa*, 11(2). Recuperado de <http://redie.uabc.mx/vol11no2/contenido-rueda3.html>
- Ruiz, J. I. (2004). *Técnicas de control y triangulación de la calidad en la investigación socioeducativa. Proyecto de Calidad Integrado*. Bilbao: Ediciones Mensajero.
- Ruiz, J. M. (2010). Evaluación del diseño de una asignatura por competencias, dentro del EEES, en la carrera de Pedagogía: Estudio de un caso real. *Revista de Educación*,

351, 435-460. Recuperado de

http://www.revistaeducacion.mec.es/re351/re351_18.pdf

Ruiz, R. (2011). La evaluación de las prácticas externas universitarias. Hacia un modelo mixto de medición de la adquisición de competencias y satisfacción de los agentes implicados. En EVALfor (Ed.), *EVALtrends 2011: Evaluar para aprender en la universidad. Experiencias innovadoras en la sistematización de la evaluación* (pp. 7-17). Madrid: Bubok Publishing.

Ruiz-Gallardo, J. R., Ruiz, E., & Ureña, N. (2013). La evaluación en la formación inicial del profesorado: Qué creemos hacer y qué perciben los alumnos. *Cultura, Ciencia y Deporte*, 8(22), 17-29. doi: 10.12800/ccd.v8i22.220

S

Sabariego, M. (2009). El proceso de investigación (Parte II). En R. Bisquerra (Coord.), *Metodología de la investigación educativa* (2ª ed.) (pp. 127-163). Madrid: La Muralla.

Sabariego, M., & Bisquerra, R. (2009a). Fundamentos metodológicos de la investigación educativa. En R. Bisquerra (Coord.), *Metodología de la investigación educativa* (2ª ed.) (pp. 19-49). Madrid: La Muralla.

Sabariego, M., & Bisquerra, R. (2009b). El proceso de investigación (Parte I). En R. Bisquerra (Coord.), *Metodología de la investigación educativa* (2ª ed.) (pp. 89-125). Madrid: La Muralla.

Sabaté, S., Marzo, M., Martínez, M., & Añaños, E. (2008). Modelos de evaluación en el marco del EEES ¿Facilitan el desarrollo de competencias? En, *Actas del V*

- Congreso Internacional Docencia Universitaria e Innovación: El Cambio en la cultura docente universitaria* [CD]. Universidad de Lleida.
- Sabirón, F., & Arraiz, A. (2013). Aprendiendo de la evaluación: Decálogo para la evaluación auténtica de competencias profesionales a través del portafolio. *Revista Iberoamericana de Evaluación Educativa*, 6(1), 135-152. Recuperado de: <http://www.rinace.net/riee/numeros/vol6-num1/art08.pdf>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144. doi: 10.1007/BF00117714
- Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535-550. doi: 10.1080/02602930903541015
- Sadler, D. R. (2013). Opening up feedback: Teaching learners to see. En S. Merry, M. Price, D. Carless, & M. Taras (Eds.), *Reconceptualising feedback in higher education. Developing dialogue with students* (pp. 54-63). London: Routledge.
- Sagasta, M. P., & Bilbatua, M. (2009). La evaluación: ¿Una herramienta o un obstáculo para el desarrollo del aprendizaje autónomo? En M. Castelló (Coord.), *La evaluación auténtica en enseñanza secundaria y universitaria: Investigación e innovación* (pp. 55-86). Barcelona: Edebé.
- Sagasta, P., Bilbatua, M., Galindez, E., & Bikuña, A. (2006). Mondragon Unibetsitatea: Propuesta de un modelo de formación a partir del perfil profesional del profesor universitario. En *Actas del IV Congreso Internacional de Docencia Universitaria e Innovación* [CD]. Universidad de Barcelona.
- Salinas, D. (2002). *¡Mañana examen! La evaluación: Entre la teoría y la práctica*. Barcelona: Graó.

- Salinas, B., & Cotillas, C. (Coords.). (2007). *La evaluación de los estudiantes en la educación superior. Apuntes de buenas prácticas*. Recuperado de <http://www3.uji.es/~betoret/Formacion/Evaluacion/Documentacion/La%20evaluacion%20estudiantes%20en%20la%20ESuperior%20UV.pdf>
- Sambell, K. (2011). *Rethinking feedback in higher education: An assessment for learning perspective*. Recuperado de <http://www.cumbria.ac.uk/Public/Education/Documents/Research/EducatorsStorehouse/Assessment/RethinkingFeedbackInHigherEducation.pdf>
- Sambell, K. (2013). Involving students in the scholarship of assessment: Student voices on the feedback agenda for change. En S. Merry, M. Price, D. Carless, & M. Taras (Eds.), *Reconceptualising feedback in higher education. Developing dialogue with students* (pp. 80-91). London: Routledge.
- Sambell, K., Gibson, M., & Montgomery, C. (2007). *Rethinking feedback: An assessment for learning perspective* (Red guides, paper 34). Recuperado de <http://www.northumbria.ac.uk/static/5007/arpdf/academy/redguide34.pdf>
- Samuelowicz, K., & Bain, J. D. (2001). Revisiting academics' beliefs about teaching and learning. *Higher Education*, 41, 299-325.
- Sánchez, J. (2011). Evaluación de los aprendizajes universitarios: Una comparación sobre sus posibilidades y limitaciones en el Espacio Europeo de Educación Superior. *Revista de Formación e Innovación Educativa Universitaria*, 4(1), 40-54. Recuperado de http://webs.uvigo.es/refiedu/Refiedu/Vol4_1/REFIEDU_4_1_4.pdf
- Sánchez, M. M. (2010). *Espacios virtuales para la evaluación de aprendizajes basados en herramientas de web semántica* (Tesis doctoral). Recuperado de <http://digitum.um.es/xmlui/handle/10201/33252>

- Sánchez, M. A; López, M. A., & Fernández, M. V. (2010). Análisis de las competencias genéricas en los nuevos títulos de grado del EEES en las universidades españolas. *Revista de Docencia Universitaria*, 8(1), 35-73. Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/145>
- Sanders, J. R., & Vogel, S. (2012). The development of standards for teacher competence in educational assessment of students. En S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 27-40). Recuperado de <http://digitalcommons.unl.edu/burosteachertraining/> (Trabajo original publicado en 1993).
- Sanmartí, N. (2007). *10 ideas clave. Evaluar para aprender*. Barcelona: Graó.
- Sanmartí, N. (2010). Aprender a evaluarse: Motor de todo aprendizaje. *Aula de Innovación Educativa*, 192, 26-29.
- Sanmartí, N., & Jorba, J (1995). Autorregulación de los procesos de aprendizaje y construcción de conocimiento. *Alambique*, 4, 59-77.
- Sans, A. (2008). *La evaluación de los aprendizajes: Construcción de instrumentos*. Recuperado de <http://www.octaedro.com/ice/pdf/DIG102.pdf>
- Sans, A. (2009). Métodos de investigación de enfoque experimental. En R. Bisquerra (Coord.), *Metodología de la investigación educativa* (2ª ed.) (pp. 167-230). Madrid: La Muralla.
- Santos, M. A. (1993). *La evaluación: Un proceso de diálogo, comprensión y mejora*. Málaga: Aljibe.
- Santos, M. A. (1999). 20 paradojas de la evaluación del alumnado en la universidad española. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 2(1), 369-399. Recuperado de dialnet.unirioja.es/descarga/articulo/2796478.pdf

- Santos, M. A. (2001). Dime cómo evalúas (en la universidad) y te diré qué tipo de profesional (y de persona) eres. *Tendencias Pedagógicas*, 6, 89-100. Recuperado de http://www.tendenciaspedagogicas.com/Articulos/2001_06_04.pdf
- Santos, M. A. (2003). *Una flecha en la diana. La evaluación como aprendizaje*. Madrid: Narcea.
- Santos, M. A. (2006). *El pato en la escuela o el valor de la diversidad*. Recuperado de <http://jportugal.wikispaces.com/file/view/El+pato+en+la+escuela.pdf>
- Santos, M. A. (2009). Almas tatuadas. Aprendizagens sobre avaliação a partir da experiência. *Sísifo. Revista de Ciências da Educação*, 9, 101-114. Recuperado de <http://dialnet.unirioja.es/servlet/articulo?codigo=3044857>
- Santos, M. A. (2010a). La formación del profesorado en las instituciones que aprenden. *Revista Interuniversitaria de Formación del Profesorado*, 68(24,4), 175-200.
- Santos, M. A. (2010b). Los peligros de la evaluación. *Cuadernos de Pedagogía*, 397, 90-93.
- Santos, M. A., & Moreno, T. (2004). ¿El momento de la metaevaluación educativa? Consideraciones sobre epistemología, método, control y finalidad. *Revista Mexicana de Investigación Educativa*, 23(9), 913-931. Recuperado de http://www.oei.es/evaluacioneducativa/momento_metaevaluacion_educativa_santos_guerra.pdf
- Saravia, M. A. (2011). *Calidad del profesorado. Un modelo de competencias académicas* (Cuadernos de docencia universitaria, 20). Recuperado de <http://www.octaedro.com/ice/cuadernos.htm>
- Saroyan, A. (2001). Evaluation of teaching for improvement: A myth or a reality? En M. Rueda, F. Díaz, & M. Díaz (Eds.), *Evaluar para comprender y mejorar la docencia en la educación superior* (pp. 151-160). México: CESU-UAM.

- Sayós, R. (2010). Evaluación continuada. En A. Parcerisa (Coord.), *Ejes para la mejora docente en la universidad* (pp. 126-132). Barcelona. Octaedro-ICE UB.
- Scriven, M. (1967). The methodology of evaluation. En R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*. (pp. 75-82). Chicago: Rand McNally.
- Scriven, M. (1968). An introduction to metaevaluation. *Educational Products Report*, 2(5), 36-38.
- Scriven, M. (1991). *Evaluation thesaurus*. Newbury Park, CA: Sage.
- Scriven, M., (2011). *Evaluating evaluations: A meta-evaluation checklist* (6th ed. rev.). Recuperado de <http://www.alnap.org/resource/8154.aspx>
- Schommer, M., Beuchat, M., & Hernández, F. (2012). Creencias epistemológicas y de aprendizaje en la formación inicial de profesores. *Anales de Psicología*, 28(2), 465-474. doi: 10.6018/analesps.28.2.125341
- Schwandt, T. A., & Halpern, E. S. (1988). *Linking auditing and metaevaluation: Enhancing quality in applied research*. Thousand Oaks, CA: Sage Publications.
- Scouller, K. M. (1996). Influence of assessment methods on students' learning approaches, perceptions, and preferences: Assignment essay versus short answer questions. *Research and Development in Higher Education*, 19(3), 776-781. Recuperado de <http://www.herdsa.org.au/wp-content/uploads/conference/1997/scoull01.pdf>
- Scouller, K. M. (1998). Influence of assessment method on students' learning approaches: Multiple choice question examination versus essay assignment. *Higher Education*, 35, 453-472. doi: 10.1023/A:1003196224280

- Serafini, F. (2000). Three paradigms of assessment: Measurement, procedure and inquiry. *The Reading Teacher*, 54(4), 384-393. Recuperado de <http://www.frankserafini.com/publications/serafini-3-parad-assess.pdf>
- Serafini, F. (2002). Reflective practice and learning. *Primary Voices*, 10(4), 2-7. Recuperado de <http://www.frankserafini.com/publications/serafini-pvoices.pdf>
- Serrano, C. (2010). Pensamiento del profesor: Un acercamiento a las creencias y concepciones sobre el proceso de enseñanza-aprendizaje en la educación superior. *Revista de Educación*, 352, 267-287. Recuperado de http://www.revistaeducacion.educacion.es/re352/re352_12.pdf
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14. doi: 10.3102/0013189X029007004
- Shulman, L. S. (2000). From Minsk to Pinsk: Why a scholarship of teaching & learning? *The Journal of Scholarship of Teaching & Learning*, 1(1), 48-53. Recuperado de <http://www20.csueastbay.edu/faculty/ofd/resources/archive/shulman.pdf>
- Siegel, S. (1975). *Estadística no paramétrica aplicada a las ciencias de la conducta* (2ª ed.). México: Trillas.
- Sime, L. (1998). Metaevaluación: Ir más allá de la evaluación para volver sobre ella. *Educación*, VII, 199-216. Recuperado de <http://revistas.pucp.edu.pe/index.php/educacion/article/viewFile/5228/5222>
- Skelton, A. (2007). International perspectives on teaching excellence in higher education. Improving knowledge and practice: Conclusion. En A. Skelton (Ed.), *International perspectives on teaching excellence in higher education. Improving knowledge and practice* (pp. 257-268). Canadá: Routledge.

Sluifmans, D., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self-, peer- and co-assessment. *Learning Environments Research*, 1(3), 293-319.
doi: 10.1023/A:1009932704458

Smitter, Y. (2006). La evaluación de los aprendizajes en una institución formadora de docentes. *Revista de Educación y Pedagogía*, 18(26), 153-166. Recuperado de <http://aprendeenlinea.udea.edu.co/revistas/index.php/revistaeyp/article/viewFile/6942/6355>

Smitter, Y., & Siso, J. M. (abril, 2007). *Evaluación estratégica del aprendizaje para mejorar la calidad educativa en la formación docente*. Trabajo presentado en la VII Reunión Nacional de Currículo. I Congreso Internacional de Calidad en Innovación en Educación Superior. Caracas. Recuperado de <http://www.cies2007.eventos.usb.ve/memorias/ponencias/12.pdf>

Southwell, D., & Morgan, W. (2009). *Leadership and the impact of academic staff development and leadership development on student learning outcomes in higher education: A review of the literature*. Recuperado del sitio de Internet del Office for Learning & Teaching (OLT): <http://www.olt.gov.au/resource-leadership-academic-staff-development-literature-review-qut-2009>

Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record*, 68, 523-540. Recuperado de <http://education.illinois.edu/circe/publications/Countenance.pdf>

Stake, R. E. (1975). To evaluate an arts program. En R. E. Stake (Ed.), *Evaluating the arts in education: A responsive approach* (pp. 13-31). Columbus-OH: Chas Merrill.

Stake, R. E. (1995). *Investigación con estudio de casos*. Madrid: Morata.

Stake, R. E. (2006). *Evaluación comprensiva evaluación basada en estándares*. Barcelona: Graò.

Stake, R. E. (2011). Program evaluation, particularly responsive evaluation. *Journal of Multidisciplinary Evaluation*, 7(15), 180-201. Recuperado de http://journals.sfu.ca/jmde/index.php/jmde_1/article/viewFile/303/298 (Trabajo original publicado en 1975).

Stiggins, R. (febrero, 1998). *Auditing the quality of classroom assessment training in teacher education programs*. Paper presentado en el Annual Meeting of the American Association of Colleges of Teacher Education. New Orleans, LA. Recuperado de http://wps.prenhall.com/wps/media/objects/1557/1594989/Volume_medialib/AUDITING.PDF

Stiggins, R. (1999). Barriers to effective student assessments. *Education Digest*, 64(6), 25-29.

Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.

Stiggins, R. (septiembre, 2005a). *Assessment FOR learning defined*. Trabajo presentado en la Conferencia internacional del Assessment Training Institute (ATI). Promoting sound assessment in every classroom. Portland, OR. Recuperado de <http://ati.pearson.com/downloads/afldefined.pdf>

Stiggins, R. (2005b). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324-328.

Stiggins, R. (2007). Assessment through the student's eyes. *Educational Leadership*, 64(8), 22-26.

Stiggins, R. (2012). Teacher training in assessment: Overcoming the neglect. En S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 27-40).

Recuperado de <http://digitalcommons.unl.edu/burosteachertraining/> (Trabajo original publicado en 1993).

Stiggins, R., Arter, J., Chappuis, J., & Chappuis, S. (2006). *Classroom assessment for student learning: Doing it right—Using it well*. Portland, OR: Pearson Assessment Training Institute.

Stiggins, R., & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory into Practice*, 44(1), 11-18. doi: 10.1207/s15430421tip4401_3

Stiggins, R., & Chappuis, J. (2008). Enhancing student learning. *District Administration*, 44(1), 42-44. Recuperado de <http://www.districtadministration.com/article/enhancing-student-learning#0>

Stiggins, R., & Conklin, N. F. (1988). *Teacher training in assessment. A final report to the Office of Educational Research and Improvement*. Recuperado de la base de datos de ERIC. (ED303439)

Stiggins, R., & Duke, D. (2008). Effective instructional leadership requires assessment leadership. *Phi Delta Kappan*, 90(4), 285-291.

Stobart, G. (2010). *Tiempos de pruebas: Usos y abusos de la evaluación*. Madrid: Ministerio de Educación-Morata.

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, 30(4), 325-341. doi: 10.1080/02602930500099102

- Stufflebeam, D. L. (1971). The relevance of the CIPP evaluation model for educational accountability. *Journal of Research and Development in Education*, 5(1), 19-25.
Recuperado de la base de datos de ERIC. (EJ048749)
- Stufflebeam, D. L. (1999). *Program metaevaluation checklist*. Recuperado de http://www.wmich.edu/evalctr/archive_checklists/program_metaeval.pdf
- Stufflebeam, D. L. (2001a). Evaluation checklists: Practical tools for guiding and judging evaluations. *American Journal of Evaluation*, 22(1), 71-79. doi: 10.1177/109821400102200107
- Stufflebeam, D. L. (2001b). The metaevaluation imperative. *American Journal of Evaluation*, 22(2), 183-209. doi: 10.1177/109821400102200204
- Stufflebeam, D. L. (2011). Meta-Evaluation. *Journal of MultiDisciplinary Evaluation*, 7(15), 99-158. Recuperado de http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/300/295 (Trabajo original publicado en 1974).
- Stufflebeam, D. L., Foley, W., Gephart, W., Guba, E., Hammond, R., Merriman, H., & Provus, M. (1971). *Educational evaluation and decision making*. Itasca, IL: Peacock.
- Stufflebeam, D. L., & Shinkfield, A. (1987). *Evaluación sistemática. Guía teórica y práctica*. Barcelona: Centro de Publicaciones del Ministerio de Educación y Ciencia-Ediciones Paidós.
- Suskie, L. (2006). *What is 'good' assessment? A Synthesis of principles of good practice*. Recuperado de <http://www2.acs.ncsu.edu/UPA/archives/assmt/resource.htm>
- Sweedler-Brown, C. O. (1992). The effect of training on the appearance bias of holistic essay graders. *Journal of Research and Development in Education*, 26(1), 24-29.
Recuperado de la base de datos de ERIC. (EJ458519)

T

- Taras, M. (2002). Using assessment for learning and learning from assessment? *Assessment & Evaluation in Higher Education*, 27, 501-510. doi: 10.1080/0260293022000020273
- Tejada, J. (2005). El trabajo por competencias en el practicum: Cómo organizarlo y evaluarlo. *Revista Electrónica de Investigación Educativa*, 7(2). Recuperado de <http://redie.uabc.mx/index.php/redie/article/view/192>
- Tejada, J. (2009). Competencias docentes. *Profesorado. Revista de currículum y formación del profesorado*, 13(2), 1-16. Recuperado de <http://www.ugr.es/~recfpro/rev132COL2.pdf>
- Tejedor, F. (Dir.). (1999). *Las estrategias utilizadas por los profesores universitarios para la evaluación del aprendizaje de los alumnos: Análisis y valoración de la situación actual, propuesta y seguimiento de experiencias de cambio dirigidas a la mejora de la calidad de la enseñanza y la profesionalización docente*. Recuperado de <http://redined.mecd.gob.es/xmlui/handle/11162/83492>
- Tejedor, F. (2003). Un modelo de evaluación del profesorado universitario. *Revista de Investigación Educativa*, 2003, 21(1), 157-182. Recuperado de <http://revistas.um.es/rie/article/view/99151>
- Tejedor, F. (2009). Evaluación del profesorado universitario: Enfoque metodológico y algunas aportaciones de la investigación. *Estudios Sobre Educación*, 16, 79-102. Recuperado de <http://dadun.unav.edu/bitstream/10171/9169/1/16%20Estudios%20Ed.pdf>
- Tenbrik, T. (1981). *Evaluación. Guía práctica para profesores*. Madrid: Narcea.

- Tigelaar, D. H., Dolmans, D. M., Wolfhagen, I. P., & Van der Vleuten, C. M. (2004). The development and validation of a framework for teaching competencies in higher education. *Higher Education*, 48(2), 253-268.
- Times Higher Education. (2013). *Times Higher Education's 2012-2013 World University Rankings*. Recuperado de <http://www.timeshighereducation.co.uk/world-university-rankings/2013-14/world-ranking>
- Tobón, S. (2010). *Formación integral y competencias. Pensamiento complejo, currículo, didáctica y evaluación*. Bogotá: Ecoe.
- Toledo, M. (2005). Competencias evaluativas y formación docente. *Revista de Orientación Educativa*, 19(36). Recuperado de <http://monica.toledos.be/competenciadocen.html>
- Toledo, M. (2006). Competencias didácticas, evaluativas y metacognitivas. *Revista de Orientación Educativa*, 20(38), 105-116. Recuperado de <http://monica.toledos.be/competencia.html>
- Tomás, M., Castro, D., & Feixas, M. (2012). Tensiones entre las funciones docente e investigadora del profesorado en la universidad. *Revista de Docencia Universitaria*, 10(1), 343-367. Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/319/pdf>
- Tonucci, F. (2010). Apuntes sobre el manifiesto. *Educar(nos)*, 49, 9-11.
- Torra, I., de Corral, I., Pérez, M., Triadó, X., Pagès, T., Valderrama, E., ... Tena, A. (2012). Identificación de competencias docentes que orienten el desarrollo de planes de formación dirigidas al profesorado universitario. *Revista de Docencia Universitaria*, 10(2), 21-56. Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/397>

- Torra, I., Màrquez, M. D., Pagès, T., Solà, P., García, R., Molina, F., ... Sangrà, A. (2013). Retos institucionales de la formación del profesorado universitario. *Revista de Docencia Universitaria*, 11(1), 285-309. Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/398>
- Torrado, M. (2009). Estudios de encuesta. En R. Bisquerra (Coord.), *Metodología de la investigación educativa* (2ª ed.) (pp. 231-257). Madrid: La Muralla.
- Torrance, H. (Ed.). (1995). *Evaluating authentic assessment: Problems and possibilities in new approaches to assessment*. Buckingham: Open University Press.
- Torre, J. C. (2004). Presentación. En J. C. Torre & E. Gil (Eds.), *Hacia una enseñanza universitaria centrada en el aprendizaje. Libro homenaje a Pedro Morales* (pp. 11-17). Madrid: Universidad Pontificia de Comillas.
- Torre, J. C. (2008). Estrategias para potenciar la autoeficacia y la autorregulación académica en los estudiantes universitarios. En L. Prieto (Coord.), *La enseñanza universitaria centrada en el aprendizaje. Estrategias útiles para el profesorado* (pp. 61-90). Barcelona: Octaedro-ICE UB.
- Trías, E. (2000). *Ética y condición humana*. Madrid: Biblioteca Nueva.
- Trillo, F. (2005). Competencias docentes y evaluación auténtica: ¿Falla el protagonista? *Perspectiva Educativa*, 45, 85-102. Recuperado de http://www.euv.cl/archivos_pdf/rev_perspectiva_educ/persp_45_1sem.pdf
- Trillo, F., & Porto, M. (1999). La percepción de los estudiantes sobre su evaluación en la Universidad. Un estudio de caso en la Facultad de Ciencias de la Educación. *Innovación Educativa*, 9, 55-75. Recuperado de https://minerva.usc.es/bitstream/10347/5189/1/pg_057-078_inneduc9.pdf

Trillo, F., & Porto, M. (2002). La evaluación de los estudiantes en el marco de la evaluación de la calidad de la universidad. *Revista de Educación*, 328, 283-301.

Tyler, R. W. (1950). *Basic principles of curriculum and instruction*. Chicago: Chicago University Press.

U

UK Professional Teaching Standards Framework (UKPTSF). (2012). *The UK Professional Standards Framework*. Recuperado de <http://www.heacademy.ac.uk/ukpsf>

Unigarro, M .A., & Rondón, M. (2005). Tareas del docente en la enseñanza flexible (el caso de UNAB Virtual). *Revista de Universidad y Sociedad del Conocimiento*, 2(1), 74-84. Recuperado de <http://www.uoc.edu/rusc/dt/esp/unigarro0405.pdf>

Universidad de Deusto (UD). (2013). *Historia y misión*. Recuperado de <http://www.deusto.es>

Universidad de Hawai'i Manoa (UH-Manoa), Assessment Office. (2013). *Assessment home. Resources: Definitions/glossary*. Recuperado de <http://manoa.hawaii.edu/assessment/resources/definitions.htm>

Ureña, N., Valles, C., & Ruiz, E. (2009). La respuesta del alumnado en los procesos de evaluación formativa y compartida. En V. M. López (Coord.), *Evaluación formativa y compartida en educación superior. Propuestas, técnicas, instrumentos y experiencias* (pp. 105-125). Madrid: Narcea.

V

- Valcárcel, M. (Dir.). (2003). *La preparación del profesorado universitario para la convergencia europea de educación superior* (Proyecto EA2003-0040). Recuperado del sitio de Internet de la Universidad de Salamanca: http://campus.usal.es/web-usal/Novedades/noticias/bolonia/informe_final.pdf
- Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, 39(2), 111-133. doi: 10.1207/s15326985ep3902_3
- Vandevoort, L. G., Amrein-Beardsley, A., & Berliner, D. C. (2004). National board certified teachers and their students' achievement. *Education Policy Analysis Archives*, 12(46), 1-117. Recuperado de la base de datos de ERIC. (EJ853513)
- Vargas, M. V. (2008). Sentido pedagógico de la evaluación. Estudio de caso en la Facultad de Ciencias de Universidad Pontificia Javeriana. *Universitas Scientiarum*, 13, 252-257. Recuperado de <http://www.scielo.org.co/pdf/unsc/v13n3/v13n3a04.pdf>
- Vázquez, J. A. (2011). Los caminos de Bolonia. *Revista de Docencia Universitaria*, 9(3), 23-38. Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/299>
- Vera, J., & Esteve, J. M. (2001). *Un examen a la cultura escolar*. Barcelona: Octaedro.
- Vera, M., Salanova, M., & Martín, B. (2010). Profesorado universitario y su bienestar laboral: La importancia del triple perfil laboral. *Electronic Journal of Research in Educational Psychology*, 8(21), 581-602. Recuperado de <http://www.investigacion-psicopedagogica.org/revista/new/ContadorArticulo.php?407>

- Verdejo, A. (2008). La evaluación del aprendizaje como uno de los fundamentos de la educación: Hacia dónde vamos. *Investigación y Evaluación Educativa*, 4(2), 7-10. Recuperado de <http://ineva.uprrp.edu/boletin/v0004n0002.pdf>
- Vicente de, M. A., & Manera, J. (2005). El análisis factorial por componentes principales. En J. P. Lèvy & J. Varela (Dir.), *Análisis multivariable para las ciencias sociales* (pp. 327-360). Madrid: Pearson Prentice.
- Vidal, J., & Quintanilla, M. A. (2000). The teaching and research relationship within institutional evaluation. *Higher Education*, 40, 220-229.
- Vilà, R., & Bisquerra, R. (2009). El análisis cuantitativo de los datos. En R. Bisquerra (Coord.), *Metodología de la investigación educativa* (2ª ed.) (pp. 259-271). Madrid: La Muralla.
- Villa, A. (2004). Convergencia europea y actualización del profesorado. En J. C. Torre & E. Gil (Eds.), *Hacia una enseñanza universitaria centrada en el aprendizaje. Libro homenaje a Pedro Morales* (pp. 271-319). Madrid: Universidad Pontificia de Comillas.
- Villa, A. (2008). La excelencia docente. *Revista de Educación, número extraordinario*, 177-212. Recuperado de http://www.revistaeducacion.mec.es/re2008_08.htm
- Villa, A., Campo, L., Arranz, S., Villa, O., & García, A. (2013). Valoración del profesorado de magisterio sobre el aprendizaje basado en competencias implantado. *Profesorado. Revista de currículum y formación del profesorado*, 17(3), 35-55. Recuperado de <http://www.ugr.es/~recfpro/rev173ART2.pdf>
- Villa, A., & García, A. (2008). *Competencias docentes en el ámbito universitario*. Manuscrito inédito, Unidad Técnica Innovación y Calidad (UTIC), Universidad de Deusto, Bilbao.

- Villa, A., & García, A. (2014). Un sistema de garantía de calidad de la docencia: Un estudio de caso. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 17(3), 65-78. doi: 10.6018/reifop.17.3.204061
- Villa, A., & Poblete, M. (Dirs.). (2007). *Aprendizaje basado en competencias: Una propuesta para la evaluación de las competencias genéricas*. Bilbao: Universidad de Deusto.
- Villa, A., & Poblete, M. (2011). Evaluación de competencias genéricas: Principios, oportunidades y limitaciones. *Bordón*, 63(1), 147-170.
- Villa, A., Carbone, R., & Troncoso, P. (2012). *Diseño e implementación de un sistema de evaluación de competencias docentes y de apoyo a su perfeccionamiento, a partir de la definición de excelencia docente, en el marco de los modelos Universidad Católica de Temuco (UCT) y Universidad Católica de la Santísima Concepción (UCSC)* (Proyecto MECESUP UCT 0805). Trabajo inédito, UCT y UCSC, Chile.
- Villar, L. M. (2009). Competencias profesionales en educación secundaria. En L. M. Villar (Coord.), *Creación de la excelencia en educación secundaria* (pp. 1-22). Madrid: Pearson Educación.
- Villardón, L. (2006). Evaluación del aprendizaje para promover el desarrollo de competencias. *Educatio Siglo XXI*, 24, 57-76. Recuperado de <http://revistas.um.es/educatio/article/viewFile/153/136>
- Villardón, L., Yániz, C., Elexpuru, I., & Achurra, C. (2013). Autonomy and competence for learning. *The European Journal of Social and Behavioral Sciences*, 4(1), 742-755. doi: 10.15405/FutureAcademy/ejsbs(2301-2218).2012.4.9
- Viso, J. R. (2010). *Enseñar y aprender por competencias. Qué son las competencias* (Vol. I). Madrid: Editorial EOS.

- Voorhees, N. K. (2001). *Measuring what matters: Competency-based learning models in higher education. New directions for institutional research*. San Francisco: Josey-Bass.
- Vries de, W. (2005). Veinte años después: Calidad, eficiencia y la educación superior. En W. De Vries (Coord.), *Calidad eficiencia y evaluación de la educación superior* (pp. 1-25). La Coruña: Netbiblo.
- Vu, T. T., & Dall'Alba, G. (2007). Students' experience of peer assessment in a professional course. *Assessment & Evaluation in Higher Education*, 32(5), 541-556. doi: 10.1080/02602930601116896

W

- Walker, M. (2013). Feedback and feedforward: Students responses and their implications. En S. Merry, M. Price, D. Carless, & M. Taras (Eds.), *Reconceptualising feedback in higher education. Developing dialogue with students* (pp. 103-112). London: Routledge.
- Watts, F., & García-Carbonell, A. (Eds.). (2006). *La evaluación compartida: Investigación multidisciplinar*. IEMA. Universidad Politécnica de Valencia.
- Westmont College. (2013). *Assessment resources: Assessment glossary*. Recuperado de http://www.westmont.edu/offices/institutional_portfolio/program_review/eeresources_assessment.html
- Weston Manor Grupo (WMG), ASKe (Assessment Standards Knowledge exchange). (2007). *Assessment standards: A manifesto for change*. Recuperado de <http://www.brookes.ac.uk/aske/documents/ManifestoLeafletNew.pdf>

- Wiggins, G. (1989). Moving to modern assessments. *Phi Delta Kappan*, 70(9), 703-713.
- Wiggins, G. (1996). Practicing what we preach in designing authentic assessments. *Educational Leadership*, 54(4), 18-25.
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3-14. doi: 10.1016/j.stueduc.2011.03.001
- Williams, R., & Brennan, J. (2003). *Collecting and using student feedback on quality and standards of learning and teaching in higher education*. Recuperado del sitio de Internet de la Higher Education Funding Council for England (HEFCE): https://www.heacademy.ac.uk/sites/default/files/id352_collecting_and_using_student_feedback_a_guide_to_good_practice.pdf
- Wilson, M. J., Diao, M. M., & Huang, L. (2015). 'I'm not here to learn how to mark someone else's stuff': An investigation of an online peer-to-peer review workshop tool. *Assessment & Evaluation in Higher Education*, 40(1), 15-32. doi: 10.1080/02602938.2014.881980
- Winniger, S. R. (2005). Using your tests to teach: Formative summative assessment. *Teaching of Psychology*, 32(3), 164-166. doi: 10.1207/s15328023top3203_7
- Wise, S. L., & Lukin, L. E. (2012). Measurement training in Nebraska teacher education programs. En S. L. Wise (Ed.), *Teacher training in measurement and assessment skills* (pp. 187-202). Recuperado de <http://digitalcommons.unl.edu/burosteachertraining/> (Trabajo original publicado en 1993).
- Wolf, D. P., & Reardon, S. F. (1996). Access to excellence through new forms of student assessment. En J. B. Baron & D. P. Wolf (Eds.), *Performance based student*

assessment: Challenges and possibilities. Ninety-fifth Yearbook of the National Society of Education (Part 1) (pp. 1-31). Chicago, IL: University of Chicago Press.

Wootton, S. (2002). Encouraging learning or measuring failure? *Teaching in Higher Education*, 7(3), 353-357. doi: 10.1080/13562510220144833

Y

Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education*, 18(3), 285-297. doi: 10.1080/13562517.2012.719154

Yániz, C. (2008). Las competencias en el currículo universitario: Implicaciones para diseñar el aprendizaje y para la formación del profesorado. *Revista de Docencia Universitaria*, 6(1). Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/59>

Yániz, C., & Villardón, L. (2006). *Planificar desde competencias para promover el aprendizaje* (Cuadernos monográficos del ICE, nº 12). Bilbao: ICE Universidad de Deusto.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45, 477-501.

Z

Zabalza, M. A. (2002). *La enseñanza universitaria. El escenario y sus protagonistas*. Madrid: Narcea.

- Zabalza, M. A. (2003). *Competencias docentes del profesorado universitario. Calidad y desarrollo profesional*. Madrid: Narcea.
- Zabalza, M. A. (2006). La convergencia como oportunidad para mejorar la docencia universitaria. *Revista Interuniversitaria de Formación del Profesorado*, 20(3), 37-69.
- Zabalza, M. A. (2008). El trabajo por competencias en la enseñanza universitaria. En I. Rodríguez (Ed.), *El nuevo perfil del profesor universitario en el EEES: Claves para la renovación metodológica* (pp. 79-114). Valladolid: Universidad Europea Miguel de Cervantes.
- Zabalza, M. A. (2011). Metodología docente. *Revista de Docencia Universitaria*, 9(3), 75-98. Recuperado de <http://red-u.net/redu/index.php/REDU/article/view/302>
- Zabalza, M. A. (2012). El estudio de las 'buenas prácticas' docentes en la enseñanza universitaria. *Revista de Docencia Universitaria*, 10(1), 17-42. Recuperado de <http://red-u.net/redu/index.php/REDU/issue/view/66>
- Zaragoza, J., Luís-Pascual, J. C., & Manrique, J. C. (2009). Experiencias de innovación en docencia universitaria: Resultados de la aplicación de sistemas de evaluación formativa. *Revista de Docencia Universitaria*, 7(4). Recuperado de <http://revistas.um.es/redu/article/view/92561>
- Zeichner, K. (2010). New epistemologies in teacher education. Rethinking the connections between campus courses and practical experience in teacher education at university. *Revista Interuniversitaria de Formación del Profesorado*, 68(24,4), 123-150.
- Zuñiga, R. (1997). La evaluación en la acción docente. En P. Apodaca & C. Lobato (Eds.), *Calidad en la universidad: Orientación y evaluación* (pp. 175-193). Barcelona: Laertes.

ESCALAS DE EVALUACIÓN DE LA CALIDAD DOCENTE

Las herramientas institucionales para la evaluación de la calidad docente analizadas en el trabajo de campo presentado en el capítulo V se encuentran detalladas en el ANEXO 09.

