

Estudio longitudinal de la influencia del aprendizaje en el rendimiento cognitivo: proyecto Normacog

Oihane Saez Atxukarro

Directora:

Dra. Natalia Ojeda del Pozo



NeuroLab



Neuropsicología de los Trastornos Médicos Severos
Neuropsychology of Severe Medical Conditions

Programa de Doctorado en Psicología

Departamento de Métodos y Fundamentos de la Psicología

Facultad de Psicología y Educación

Universidad de Deusto

 **Deusto**

Universidad de Deusto
Deustuko Unibertsitatea
University of Deusto



Programa de Doctorado en Psicología

Departamento de Métodos y Fundamentos de la Psicología

Facultad de Psicología y Educación

**Estudio longitudinal de la influencia del aprendizaje en
el rendimiento cognitivo: proyecto Normacog**

Tesis doctoral presentada por Oihane Sáez Atxukarro

Para la obtención del grado de Doctora por la Universidad de Deusto

Estudiante PhD

Oihane Saez Atxukarro

Directora

Dra. Natalia Ojeda del Pozo

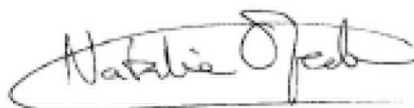
Esta tesis se ha llevado a cabo en el equipo de investigación Neuropsicología de los Trastornos Médicos Severos, Departamento de Métodos y Fundamentos de la Psicología, Facultad de Psicología y Educación, Universidad de Deusto. Este grupo ha sido calificado con la máxima acreditación de investigación otorgada por el Gobierno Vasco (Categoría Tipo A-Excelencia).

El presente trabajo ha sido financiado por el Departamento de Salud del Gobierno Vasco [2011111102 a la Dra. Natalia Ojeda del Pozo], por el Proyecto de Investigación Básica y Aplicada del Gobierno Vasco [PI_2014_1_43 al Dr. Javier Peña Lasa], por la Universidad de Deusto [FPI-UD a Oihane Sáez Atxukarro] y por la Fundación Vicente Mendieta y Lambarri [“Ayudas a la Investigación 2019” a Oihane Sáez Atxukarro].

El equipo quiere agradecer al Dr. David Schretlen, director de la División de Psiquiatría Médica, de Johns Hopkins Medical Institutions en Baltimore (Estados Unidos) por su orientación y participación en el desarrollo y puesta en marcha de este proyecto apoyando así la importancia que los procesos de normalización y baremación de instrumentos neuropsicológicos tiene en diferentes regiones del mundo.

La Dr. Natalia Ojeda del Pozo, investigadora principal del equipo de investigación Neuropsicología de los Trastornos Médicos Severos, directora y profesora titular del Departamento de Métodos y Fundamentos de la Psicología de la Universidad de Deusto, y directora de la presente tesis certifica que la presente tesis titulada “Estudio longitudinal de la influencia del aprendizaje en el rendimiento cognitivo: proyecto Normacog” representa un estudio de investigación original, presentado por Oihane Sáez Atxukarro para obtener el grado de Doctora en Psicología.

Directora

A handwritten signature in black ink, reading "Natalia Ojeda", written in a cursive style. The signature is enclosed within a faint, light-colored oval border.

Dra. Natalia Ojeda del Pozo

Bilbao, junio 2021

Agradecimientos,

A mi directora, Natalia, gracias por la confianza depositada en mi desde el primer momento, y por haberme dado la oportunidad de realizar esta tesis doctoral. Gracias por guiarme en este proceso y dedicarme tu tiempo siempre que lo he necesitado. Nuestro primer trabajo de investigación juntas, también en el marco del Proyecto Normacog, comenzó en 2012. La idea de hacer el doctorado me atraía enormemente, y cuando decidí posponerla para hacer mi segundo máster, me dijiste “...y si piensas algún en día en retomarlo, ya sabes dónde estoy”. Y como no podía ser de otra forma, dos años después, volví. Muchísimas gracias por abrirme la puerta a la investigación, a Neurolab y a esta experiencia.

A Javi, gracias por haber estado en cada uno de los pasos necesarios para que esta tesis fuera posible. Gracias por tu tiempo, y por guiarme en este proceso. Y por supuesto, gracias por resolverme todas y cada una de las preguntas de estadística que te he hecho a lo largo de estos años, y por tu infinita paciencia y comprensión.

A Naroa, gracias por tu ayuda y por cada una de las correcciones y sugerencias, que han mejorado los *papers* que hoy componen esta tesis. Gracias por preocuparte por mi y ofrecerme tu ayuda para llevar a cabo este proyecto.

A los tres, gracias por ser un ejemplo de trabajo duro, por dejarme aprender de vuestra experiencia, y por transmitirme el cariño que hoy le tengo a la investigación.

A todos los participantes de este estudio, GRACIAS. Sin vuestra colaboración desinteresada ni esta tesis ni el Proyecto Normacog hubieran sido posibles. Muchísimas gracias, de corazón, por contribuir tanto en una investigación que no siempre entendíais 100%. Ojalá pudiera nombraros a cada uno de los 997 que participasteis, porque cada uno de vosotros es importante en este trabajo.

A todos los investigadores y evaluadores, y a los diferentes centros que han participado en el proyecto, gracias; en especial a Rocío, por ser un pilar fundamental en el Proyecto Normacog, por tu apoyo incondicional, por sacar tiempo entresemana para revisar *papers* y por las noches y los fines de semana para revisar datos juntas.

A la Universidad de Deusto, mi *alma mater*, gracias por permitirme realizar aquí mi tesis doctoral, y por la concesión de una beca predoctoral que ha permitido financiar los tres primeros años de mi tesis doctoral. A la Fundación Vicente de Mendieta y Lambarri, gracias por la concesión de las “Ayudas a la Investigación 2019”, que permitieron financiar el cuarto año de mi tesis.

A Sophia Amor, autora de la ilustración que aparece en la portada, gracias por atender con tantísimo mimo a cada uno de los detalles que te he ido proponiendo, has conseguido que esta ilustración sea realmente representativa de mis evaluaciones.

A Acebo y Leire, gracias por preocuparos y preguntarme siempre por las evaluaciones y los *papers*, y por tranquilizarme y animarme en los momentos menos agradables de este proceso.

A mis Neurogirls, gracias por miles de razones. Por ser ejemplo de un trabajo bien hecho, sois las mejores investigadoras que he conocido, y habéis sido un ejemplo enorme para mí. Por ayudarme a que esta tesis sea posible. Por hacerme llorar de risa, y por hacerme reír en los momentos malos. Por constituir esta red de apoyo, que se ha

convertido en la red de seguridad que no me ha permitido caer en ningún momento. Habéis sido muchísimo más que compañeras. Por todo ello, gracias.

A ama y a aita, gracias. Por ser ejemplo de trabajo duro, me habéis inculcado desde pequeña que el esfuerzo es imprescindible cuando quieres lograr objetivos. Por permitirme estudiar absolutamente todo lo que he querido (carrera, másters, idiomas, música...), sé que ha sido un esfuerzo para vosotros, y sé que lo haríais mil veces más. Por no decirme nunca que no cuando os pedía un libro. Y por supuesto, algo imprescindible para una futura investigadora, por dar respuesta a mi insaciable curiosidad, buscando información para responder siempre a todas mis preguntas, con lo que eso suponía en los 90 sin acceso a internet.

A Jou, gracias. Por el interés que has mostrado siempre en lo que yo hacía, por aguantar mis charlas interminables, y por ofrecerte a participar en todas las investigaciones del lab. He intentado ser un buen ejemplo para ti, aunque has demostrado no necesitarlo; gracias por esforzarte en pasar tiempo conmigo siempre y por enseñarme tanto desde pequeñajo. Qué suerte tengo de ser tu *tata*.

A mi familia, gracias. Gracias amama, por preocuparte por mi cada día, y preguntarme cada tarde durante estos cinco años “¿ya has salido de clase?”. A Mary Stream. Yo también lo siento por no haberte dedicado tiempo los últimos meses. Nunca nada me ha parecido tan difícil como tener que decirte yogur.

A Carlos, gracias. Por animarme siempre a superarme; tu confianza en mis logros me ha ayudado a esforzarme para conseguirlos. Por haber estado siempre cerca, incluso cuando estabas lejos. Por haberte adaptado a mi ciudad, a mis horarios y a mi calendario. Gracias por ayudarme a relativizar, por ser calma en la tormenta y brújula cuando necesito reorientarme. Por cuidarme y cocinarme cosas ricas. Gracias por estar siempre, porque contigo todo parece más fácil.

A la gente que se ha preocupado por mi durante este tiempo, gracias. Monika, gracias por preocuparte por mi todos y cada uno de los días de los últimos años y por cuidarme tantísimo durante el confinamiento. A mis amigos, que se han preocupado por seguir con interés los pasos que iba dando, gracias.

*A todos y cada uno de los participantes de esta tesis,
en especial a mis padres, al pequeño Jou y a Carlos.*

Nada en la vida debe ser temido, solamente comprendido. Ahora es el momento de comprender más, para temer menos.

Marie Curie

Un país sin investigación es un país sin desarrollo.

Margarita Salas

Contenidos

Prólogo	xvii
Glosario de abreviaturas	xix
1. Resumen	23
2. Introducción	35
2.1. Evaluación neuropsicológica	35
2.1.1. Factores que influyen en la evaluación neuropsicológica.	37
2.1.1.1. Variables sociodemográficas.....	37
2.1.1.1.1. Edad	37
2.1.1.1.2. Educación.....	38
2.1.1.1.3. Sexo.....	38
2.1.1.1.4. Cultura, raza y etnia	39
2.1.1.2. Variables que influyen en las evaluaciones longitudinales sucesivas. ...	40
2.1.1.2.1. Efecto de la práctica.....	40
2.1.1.2.2. Fiabilidad y error de medida del instrumento.	41
2.1.1.2.3. Regresión a la media.....	41
2.2. Métodos de comparación de las puntuaciones del rendimiento.....	42
2.2.1. Comparación normativa.	42
2.2.1.1. Normas basadas en frecuencias.	43
2.2.1.2. Normas basadas en regresión.	44
2.2.2. Comparación individual	44
2.2.2.1. Pruebas con formas paralelas.	46
2.2.2.2. Índices de cambio fiable.....	47
2.2.2.2.1. Puntuaciones de discrepancia.	48
2.2.2.2.2. Índice de Desviación Estándar (IDE).	48

2.2.2.2.3. Índice de Cambio Fiable (ICF)	49
2.2.2.2.4. Índice de Cambio Fiable + efecto de la práctica (ICF+EP).....	49
2.2.2.2.5. Fórmulas basadas en regresión estandarizadas.....	50
2.2.2.2.6. Ventajas y desventajas de los diferentes índices de cambio fiable..	50
2.2.2.2.7. Interpretación de las puntuaciones z.....	51
2.3. Normalización en población española	52
2.4. Normacog.....	53
2.4.1. Protocolo Normacog.....	54
2.4.1.1. Datos normativos disponibles.....	56
2.4.1.2. HVLТ-R.....	56
2.4.1.3. BVMT-R.....	59
2.4.1.4. UD Interferencia	61
3. Acercamiento al presente estudio, objetivos e hipótesis.....	65
3.1. Estudio I.....	65
3.2. Estudio II.....	66
3.3. Estudio III	66
3.4. Estudio IV	68
4. Métodos	73
4.1. Muestra de los estudios.....	73
4.1.1 Estudio I.....	74
4.1.2 Estudios II y III.	74
4.1.3 Estudio IV.	75
4.2 Medidas cognitivas / batería de pruebas	75
4.2.1 Estudio I.....	75
4.2.2 Estudio II.....	76

4.2.3 Estudio III.....	76
4.2.4 Estudio IV.	77
4.3 Análisis estadísticos.....	78
4.3.1 Análisis sociodemográficos.	78
4.3.2 Normalización y estandarización.	79
4.3.2.1 Estudios I, II y III.	79
4.3.3 Análisis de fiabilidad y validez.	79
4.3.3.1 Estudio III.	79
4.3.4 Equivalencia entre versiones.....	80
4.3.4.1 Estudio I.....	80
4.3.5 Obtención de índices de cambio fiable.	80
4.3.5.1 Estudio IV.....	80
5. Resultados	85
5.1 Estudio I.....	87
<i>“Test de Aprendizaje Verbal de Hopkins-Revisado: normalización y estandarización de la prueba en población española”</i>	
5.2 Estudio II.....	97
<i>“Test Breve de Memoria Visuoespacial-Revisado: normalización y estandarización de la prueba en población española”</i>	
5.3 Estudio III	107
<i>“Test UD Interferencia: creación y validación de un nuevo instrumento de resistencia a la interferencia. Normalización y estandarización en población española”</i>	
5.4 Estudio IV	121
<i>“Índices de cambio fiable para la evaluación del cambio en 16 pruebas de evaluación neuropsicológica en seis momentos temporales diferentes”</i>	
6. Discusión.....	171

7. Conclusiones.....	183
8. Referencias	189
Anexo I.....	213

Prólogo

La presente tesis, presentada para la obtención del grado de Doctora por la Universidad de Deusto, es el resultado de cuatro estudios realizados en el equipo de investigación Neuropsicología de los Trastornos Médicos Severos, Departamento de Métodos y Fundamentos de la Psicología, Facultad de Psicología y Educación, Universidad de Deusto. Como resultado del trabajo llevado a cabo, se han publicado tres de los siguientes estudios en revistas revisadas por pares.

Estudio I

Sáez-Atxukarro, O., del Pino, R., Peña, J., Schretlen, D. J., Ibarretxe-Bilbao, N. y Ojeda, N. (2021). Test de aprendizaje verbal de Hopkins revisado: normalización y estandarización de la prueba en población española. *Revista de Neurología*, 72(02), 35-42. <https://doi.org/10.33588/rn.7202.2020412> [IF: 0.870, Q4 neurología clínica].

Estudio II

Sáez-Atxukarro, O., Del Pino, R., Peña, J., Schretlen, D. J., Ibarretxe-Bilbao, N. y Ojeda, N. (2021). Test Breve de Memoria Visuoespacial-Revisado: normalización y estandarización de la prueba en población española. *Revista de Neurología*, 72(09), 299–306. <https://doi.org/10.33588/rn.7209.2020527> [IF: 0.870, Q4 neurología clínica].

Estudio III

Sáez-Atxukarro, O., Del Pino, R., Peña, J., Schretlen, D. J., Ibarretxe-Bilbao, N. y Ojeda, N. (2021). Test UD Interferencia: creación y validación de un nuevo instrumento de resistencia a la interferencia. Normalización y estandarización en población española. *Neurología*. <https://doi.org/10.1016/j.nrl.2021.01.014>. [IF: 3.109, Q3 neurología clínica].

Estudio IV

Índices de cambio fiable para la evaluación del cambio en 16 pruebas de evaluación neuropsicológica en seis momentos temporales diferentes (*en preparación*).

Glosario de abreviaturas

AVDL = Actividades de la Vida Diaria de Lawton

BVMT-R = Test Breve de Memoria Visuoespacial-Revisado / *Brief Visuospatial*

Memory Test-Revised

BTA = Test Breve de Atención / *Brief Test of Attention*

CNNS = Sistema Normativo de Calibrado Neuropsicológico / *Calibrated*

Neuropsychological Normative System

DE = Desviación Estándar

DE_{norm} = Desviación Estándar de la muestra normativa

EED = Error estándar de las diferencias entre medidas

EEM = Error estándar de medida

FCT = Figura Compleja de Taylor

GDS = Escala de depresión geriátrica Yesavage / *Geriatric Depression Scale Yesavage*

GPT = *Grooved Pegboard Test*

HVLT-R = Test de Aprendizaje Verbal de Hopkins-Revisado / *Hopkins Verbal*

Learning Test-Revised

ICF = Índice de Cambio Fiable

ICF+PE = Índice de Cambio Fiable + efecto de la práctica

IDE = Índice de Desviación Estándar

INE = Instituto Nacional de Estadística

M-WCST = Test de Clasificación de Tarjetas de Wisconsin - Modificado / *Modified*

Wisconsin Card Sorting Test

MANOVA = Análisis Multivariante de la Varianza / *Multivariate analysis of variance*

MoCA = Evaluación Cognitiva de Montreal / *Montreal Cognitive Assessment*

Pc = Percentiles

PD = Puntuaciones directas

PE = Puntuaciones escalares

PEae = Puntuaciones escalares ajustadas para cada rango de edad

PROLEC-PP = Subtest de lectura de Pseudopalabras del PROLEC-R

TAP = Test de Acentuación de Palabras

TCPS = Test de Comparación Perceptual de Salthouse

TDB = Test de Denominación de Boston

TICS = Entrevista Telefónica del Estado Cognitivo / *Telephone Interview for Cognitive*

Status

TMP = Test de Memoria Prospectiva

TMT-A = *Trail Making Test-A*

TMT-B = *Trail Making Test-B*

I. Resumen

1. Resumen

1.1. Resumen

La realización de una evaluación neuropsicológica de calidad implica tener en cuenta la influencia de variables sociodemográficas como la edad de la persona evaluada, su nivel educativo, su sexo y su cultura o lengua, ya que estos factores han demostrado ser características relevantes e influyentes en el rendimiento cognitivo. Seleccionar un conjunto de datos normativos apropiados es imprescindible para una más fiable interpretación de los datos obtenidos. Además, tanto en el entorno clínico como en el entorno de la investigación es habitual la realización de evaluaciones sucesivas, por ejemplo, para hacer un seguimiento a lo largo del tiempo con el objetivo de valorar el grado deterioro asociado a determinadas enfermedades, o para evaluar el resultado de una intervención. Las evaluaciones sucesivas, sin embargo, están influidas por el efecto de la práctica, por el error de medida inherente a los test y por el efecto de regresión a la media, por lo que se hace patente la necesidad de obtención de métodos fiables que permitan la comparación individual. Los índices de cambio fiables son procedimientos estadísticos que tratan de tener en cuenta las variables que pudieran estar afectando a las evaluaciones sucesivas, proporcionando un rango de distribución de las puntuaciones de cambio esperables cuando no existe un cambio real en el nivel cognitivo del paciente. Poder contar con datos que nos permitan no sólo tener en cuenta el cambio esperado, sino evaluar la magnitud de este cambio, se hace especialmente relevante.

Esta tesis está compuesta por cuatro estudios que abordan estas cuestiones y tratan de plantear posibles soluciones. El *primer estudio* proporciona normas de corrección (basadas en N=700) adecuadas a las características sociodemográficas de la población española adulta para el Test de Aprendizaje Verbal de Hopkins-Revisado (HVLTR), basándonos en dos de sus seis versiones, que permiten corregir esta prueba teniendo en

cuenta la edad y la educación de la persona evaluada. De la misma forma que el anterior, el *segundo estudio* proporciona normas de corrección (basadas en N=905) para el Test Breve de Memoria Visuoespacial-Revisado (BVMT-R), basándonos en una de sus versiones. Ambos son test de evaluación del aprendizaje y la memoria, que cuentan con versiones alternativas que contribuyen a reducir el efecto del aprendizaje. El *tercer estudio* presenta una versión del test de Stroop, que hemos denominado test UD Interferencia, ofrece sus propiedades psicométricas, y proporciona datos normativos (basadas en N=905) adecuados a las características sociodemográficas de la población española. El *cuarto estudio* tuvo como objetivo proporcionar índices de cambio fiable para 16 pruebas de evaluación neuropsicológica, mediante cinco métodos diferentes, que permitirán una adecuada interpretación del cambio entre evaluaciones longitudinales sucesivas, teniendo en cuenta variables como la fiabilidad de la prueba o el efecto de la práctica, habitual en reevaluaciones.

Los resultados mostraron que existe un efecto significativo de la edad y el nivel educativo sobre las variables analizadas en HVLT-R, BVMT-R y UD Interferencia; como era de esperar, a mayor edad y menor nivel educativo, el rendimiento en estas pruebas fue menor. El sexo, sin embargo, sólo tuvo una ligera influencia en algunas de las variables analizadas. Además, se ha comprobado que el test UD Interferencia tiene una buena fiabilidad y validez. Por último, se observó una mejora significativa en la puntuación de la segunda evaluación cuando las pruebas fueron readministradas (1 mes, 3 meses, 6 meses, 9 meses, 12 meses o 24 meses tras la evaluación basal), especialmente en pruebas relacionadas con la memoria, y en menor medida en el grupo evaluado tras 24 meses.

El objetivo de esta tesis es aumentar la disponibilidad de herramientas para la adecuada evaluación neuropsicológica en población española reduciendo posibles limitaciones derivadas de procesos previos de baremación y normalización, y,

esperablemente facilitar la labor asistencial de los neuropsicólogos clínicos e investigadores. Teniendo esto en cuenta, se proporcionan, por un lado, baremos normativos y estandarizados para tres pruebas de evaluación neuropsicológica, contribuyendo así a aumentar la disponibilidad de baremos adaptados a población española. Por otro lado, se proporcionan cinco índices de cambio fiable para 16 pruebas de evaluación, contribuyendo así con una herramienta que permite una correcta interpretación del cambio observado entre evaluaciones sucesivas y nos proporciona información sobre la magnitud de este cambio.

Palabras clave: UD Interferencia. HVLT-R. BVMT-R. Normalización. Estandarización. Índices de cambio fiable. Efecto de la práctica. Población española. Test neuropsicológicos. Normacog.

1.2. Abstract

A quality neuropsychological evaluation should take into account the influence of sociodemographic variables such as the age of the person being evaluated, his or her educational level, gender, and culture or language, since these factors have been shown to be relevant and influential characteristics in cognitive performance. Selecting an appropriate normative data set is essential for a more reliable interpretation of the obtained data. In addition, repeated assessments are a relatively common occurrence both in clinical environment and in research environment, for example, to follow the degree of cognitive decline associated with certain diseases or to evaluate the outcome of an intervention. However, successive evaluations are biased by the practice effect, by the error of measurement inherent in the tests and by the effect of regression to the mean, so obtaining reliable methods that allow individual comparison is needed. Reliable change indices are statistical procedures that attempt to take into account the variables that could be affecting successive evaluations, providing a range of distribution of expected scores of change in the absence of real change in patient's cognitive level. Data that allows us not only to take into account the expected change, but also to evaluate the magnitude of the change, becomes especially relevant.

This thesis is composed of four studies that address these issues and attempt to propose possible solutions. The *first study* provides normative data (based on N=700) for the Spanish adult population for the Hopkins Verbal Learning Test-Revised (HVLTR), based on two of its six versions, which allows the correction taking into account the age and level of education of the person being assessed. In the same way, the *second study* provides normative data (based on N=905) for the Brief Visuospatial Memory Test - Revised (BVMT-R), based on one of its versions. Both test are measures of learning and memory, which have alternative versions that help to reduce the learning effect. The *third*

study presents a new version of the Stroop test, which we have named the UD Interference test, as well as its psychometric properties, and provides normative data (based on N=905) taking into account the Spanish sociodemographic characteristics. The *fourth study* aimed to provide reliable change indices for 16 neuropsychological assessment tests, by five methods, which will allow an adequate interpretation of the change occurred between successive longitudinal evaluations, taking into account variables such as test reliability or practice effect, which is common in reevaluations.

Results showed that UD Interference test has good reliability and validity. On the other hand, a significant effect of age and educational level on the analyzed variables of the UD Interference, the HVLТ-R and the BVMT-R was observed; as expected, the older and the less educated obtained lower performance on these tests. Sex, however, had only a slight influence on some of the analyzed variables. Finally, a significant improvement was observed in the score of the second evaluation when the tests were re-administered (1 month, 3 months, 6 months, 9 months, 12 months or 24 months after the baseline assessment), especially in memory related tests, and to a lesser extent in the group evaluated after 24 months.

The aim of this thesis is to increase the availability of tools for the adequate neuropsychological assessment in Spanish population, reducing possible limitations derived from previous standardization and normalization processes, and hopefully, facilitating the assistance work of clinical neuropsychologists and researchers. Bearing this in mind, we provide, on the one hand, normative and standardized data for three neuropsychological assessment tests, contributing to increase the availability of scales adapted to the Spanish population. On the other hand, five reliable change indices for 16 assessment tests are provided, contributing with a tool that allows a correct interpretation

of the change observed between successive assessments and provides information about the magnitude of this change.

Key words: UD Interference. HVLT-R. BVMT-R. Normalization. Standardization. Reliable change indices. Practice effects. Spanish population. Neuropsychological test. Normacog.

1.3. Laburpena

Kalitatezko ebaluazio neuropsikologiko bat egiteak aldagai soziodemografikoen eragina kontuan hartzea suposatzen du, hala nola ebaluatutako pertsonaren adina, hezkuntza maila, sexua eta kultura edo hizkuntza, errendimendu kognitiboan faktore horiek ezaugarri garrantzitsuak eta eragin handikoak direla frogatu baitute. Datu normatibo multzo egokia hautatzea ezinbestekoa da lortutako datuen interpretazioa fidagarriagoa izateko. Gainera, bai ingurune klinikoan eta baita ikerketa-ingurunean ere, segidako ebaluazioak ohikoak dira, adibidez, denboran zehar jarraipena egiteko zenbait gaixotasunekin lotutako narriadura-maila baloratzeko, edo esku-hartze baten emaitza ebaluatzeko. Segidako ebaluazioak, praktikaren efektuengatik, probek berezko duten neurketa-erroreengatik eta batez bestekoaren erregresio-efektuengatik daude eraginda, horregatik, aldaketa indibiduala neurtzeko metodo fidagarriak behar dira. Aldaketa indize fidagarriak segidako ebaluazioetan eragina izan dezaketen aldagaiak kontuan hartzen saiatzen diren prozedura estatistikoak dira, pazientearen maila kognitiboan benetako aldaketarik ez dagoenean espero diren aldaketen puntuazioen banaketa sorta eskainiz. Bereziki garrantzitsua da espero diren aldaketez gain, aldaketa hauen magnitudea ebaluatzea ahalbidetzen duten datuak edukitzea.

Tesi hau gai honei heldu eta irtenbide posibleak proposatzen saiatzen diren lau ikerketek osatzen dute. *Lehenengo ikerketak* Espainiako biztanleria helduaren ezaugarri soziodemografikoetara egokitutako datuak eskaintzen ditu (N=700 izanik) Hopkinsen Ikaskuntza Berbalerako Testa - Berrikusiarentzat (HVLT-R), bere sei bertsioetatik bitan oinarrituta, ebaluatutako pertsonaren adina eta hezkuntza kontuan hartuta proba hori zuzentzea ahalbidetzen dutenak. Aurrekoaren modu berean, *bigarren ikerketak* datu normatiboak eskaintzen ditu (N=905 izanik), Memoria Bisuoespazialerako Test Laburra - Berrikusiarentzat (BVMT-R), bere bertsioetatik baten oinarrituta. Biak ikaskuntza eta

memoria ebaluatzeko probak dira, ikaskuntzaren eragina murrizten laguntzen duten ordeko bertsioak dituztenak. *Hirugarren ikerketak* UD Interferentzia izendatu dugun testa aurkezten du, Stroop testaren bertsio bat, bere propietate psikometrikoak eskaintzen ditu eta Espainiako biztanleriaren ezaugarri soziodemografikoetara egokitutako datu normatiboak eskaintzen ditu (N=905 izanik). *Laugarren ikerketak* 16 ebaluazio froga neuropsikologikorako aldaketa indize fidagarriak eskaintzea zuen helburutzat, bost metodo desberdin erabiliz, segidako ebaluazio longitudinalen arteko aldaketaren interpretazio egokia ahalbidetuko dutenak, probaren fidagarritasuna edo praktikaren efektua bezalako aldagaiak kontuan hartuz, ohikoak berrebaluazioetan.

Emaitzek UD Interferentziak fidagarritasun eta baliozkotasun ona duela erakutsi zuten. Bestalde, adinaren eta hezkuntza mailaren eragin nabarmena ikusi zen UD Interferentzia, HVLT-R eta BVMT-R testetan aztertutako aldagaietan; espero bezala, zenbat eta adina altuagoa izan eta hezkuntza maila baxuagoa izan, probetan izandako errendimendua baxuagoa izan zen. Sexuak, aldiz, soilik eragin txikia izan zuen aztertutako aldagai batzuetan. Azkenik, segidako ebaluazioetan, bigarren ebaluazioko puntuazioan hobekuntza nabarmena ikusi zen (lehenengo ebaluazioa eta hilabete 1, 3 hilabete, 6 hilabete, 9 hilabete, 12 hilabete edo 24 hilabete pasa ondoren), batez ere memoriarekin lotutako probetan, eta neurri txikiagoan 24 hilabete igaro ondoren ebaluatutako taldean.

Tesi honen helburua Espainiako biztanleriarentzat egokitutako ebaluazio tresnen eskuragarritasuna areagotzea da, aurreko ebaluazio eta normalizazio prozesuetatik eratorritako muga posibleak murriztuz, eta, neuropsikologo kliniko eta ikertzaileen arreta lana erraztuz. Hori kontuan hartuta, alde batetik, datu normatiboak eta estandarizatuak ematen dira hiru ebaluazio probetarako, horrela, Espainiako populazioari egokitutako baremoen eskuragarritasuna handitzen lagunduz. Bestalde, aldaketa fidagarria neurtzeko

bost indize ematen dira 16 ebaluazio-probetarako, horrela, segidako ebaluazioen artean ikusitako aldaketaren interpretazio zuzena ahalbidetu eta aldaketa horren magnitudeari buruzko informazioa eskaintzen digun tresna eskaintzen da.

Hitz gakoak: UD Interferentzia. HVLТ-R. BVMT-R. Normalizazioa. Estandarizazioa. Aldaketa indize fidagarriak. Praktikaren efektua. Espainiako biztanleria. Proba neuropsikologikoak. Normacog.

II. Introducción

2. Introducción

2.1. Evaluación neuropsicológica

La neuropsicología cuenta con diferentes métodos de evaluación, algunos compartidos con otras ciencias, como los métodos anatómicos, las técnicas de neuroimagen (anatómica y funcional), las técnicas de registro (mediante electroencefalograma o potenciales evocados) y las técnicas psicofísicas, y un método propio y específico de la neuropsicología, la evaluación neuropsicológica, realizada mediante la utilización de las diferentes pruebas, escalas o test de evaluación (Portellano, 2005) complementadas habitualmente por una entrevista previa y por la historia clínica del paciente (Villa Rodríguez et al., 2017).

La evaluación neuropsicológica, que debe ser amplia y exhaustiva, ayuda a establecer perfiles neuropsicológicos y a describir alteraciones cognitivas, emocionales y conductuales (Arnett y Forn Frias, 2007). Si bien las pruebas de cribado ofrecen una información general, la evaluación neuropsicológica proporciona una valoración completa de los diferentes dominios cognitivos; habitualmente se evalúan el lenguaje, la atención, la percepción visuoespacial, las funciones ejecutivas y el aprendizaje y la memoria verbal y no verbal entre otros (Mitrushina et al., 2005; Zucchella et al., 2018). El uso de diferentes pruebas neuropsicológicas nos permite poder detectar disfunciones cognitivas en etapas tempranas y mejorar la precisión de un posible diagnóstico. Estas pruebas constituyen el método más específico para la evaluación y son, con frecuencia, más sensibles que otras técnicas neurológicas, neurofisiológicas o de neuroimagen a los efectos del daño cerebral (Portellano, 2005), ya que algunas lesiones no muestran ningún signo visible en la neuroimagen, pero tienen déficits cognitivos tan notables que les impiden un funcionamiento normal similar al previo a la lesión o enfermedad; en muchas

ocasiones el alcance del trastorno solamente queda patente mediante los resultados de las pruebas neuropsicológicas (Kolb y Whishaw, 2015).

Los objetivos de la evaluación neuropsicológica se pueden resumir, según diferentes autores, en los siguientes (Gil, 2019; Kolb y Whishaw, 2015; Lezak et al., 2012; Portellano, 2005):

- Estudio de la capacidad cognitiva de una persona y valoración del impacto de las lesiones cerebrales sobre las distintas funciones cognitivas, emocionales, y la personalidad.

- Apoyo en el diagnóstico de patologías neurológicas o psiquiátricas. En ocasiones, otros estudios de diagnóstico no pueden, por si mismos, identificar las consecuencias de una lesión o primeros síntomas de una enfermedad degenerativa.

- Diseño y preparación de programas de atención, tratamiento y rehabilitación adaptados a las necesidades de cada persona teniendo en cuenta las áreas más afectadas y las mejor preservadas.

- Valoración de la evolución y recuperación del paciente mediante seguimientos longitudinales, tanto cuando ha seguido un programa de rehabilitación cognitiva como cuando no ha recibido ningún tipo de atención. El seguimiento longitudinal permite, además, valorar la eficacia que los tratamientos de rehabilitación tienen.

Existen diferentes razones para realizar una evaluación, como por ejemplo determinar la capacidad y competencia de una persona para realizar actividades como retomar su trabajo previo o gestionar asuntos financieros, determinar la presencia o ausencia de déficits y las fortalezas y debilidades, obtener información para la realización de un diagnóstico, clarificar los efectos de una determinada enfermedad, determinar los apoyos necesarios o proporcionar información sobre la evolución esperable a largo plazo

(Strauss et al., 2006; Zucchella et al., 2018). Dado que los resultados de un examen neuropsicológico pueden ayudar a los distintos profesionales de la salud a establecer las fortalezas y debilidades funcionales, y con ello, las diferentes estrategias terapéuticas, se considera la evaluación un componente esencial en el diagnóstico, la rehabilitación, y el cuidado de los pacientes (Arnett y Forn Frias, 2007; Mitrushina et al., 2005).

Una batería de pruebas neuropsicológicas, idealmente, debe contar con medidas bien validadas, fiables, normalizadas y estandarizadas (Mitrushina et al., 2005). Teniendo en cuenta que algunas de las pruebas se utilizan extensamente en diferentes países, y con el objetivo de utilizar un lenguaje común en neuropsicología, se ha dedicado un gran esfuerzo a la normalización y estandarización de pruebas de evaluación (Ardila y Rosselli, 2007).

2.1.1. Factores que influyen en la evaluación neuropsicológica.

2.1.1.1. Variables sociodemográficas.

La realización de una evaluación neuropsicológica de calidad implica tener en cuenta la influencia de variables sociodemográficas como la edad de la persona evaluada, su nivel educativo y lengua, su sexo, y su cultura o etnia (Lezak et al., 2012; Mitrushina et al., 2005; Urbina, 2004).

2.1.1.1.1. Edad

La edad es un factor que influye significativamente en el rendimiento ante diferentes pruebas neuropsicológicas, asociándose una mayor edad a un rendimiento menor (Salthouse, 2019). Este declive está asociado a diferentes causas (Lezak et al., 2012; Zec, 1995): 1) Cambios en el cerebro, registrados a partir de los 40-50 años; 2) Envejecimiento normal, asociado a cambios neurobiológicos que ocurren con la edad y contribuyen al deterioro de las habilidades cognitivas; 3) Enfermedades que aumentan con la edad y hábitos de vida asociados al envejecimiento que pueden influir y

comprometer el adecuado funcionamiento cognitivo (como la hipertensión, diabetes, o los malos hábitos nutricionales); 4) Enfermedades cerebrovasculares y degenerativas. La reducción en el uso de algunas habilidades contribuye igualmente a su deterioro (Zec, 1995). Además del deterioro cognitivo asociado a la edad, se deben tener en cuenta las características socioculturales de algunos grupos etarios. En el caso de la población española, las personas más mayores han vivido en un contexto cultural e histórico que ha contribuido a que el rendimiento cognitivo sea menor al observado en otras culturas (Ojeda, Aretouli, et al., 2016).

2.1.1.1.2. Educación

La educación se asocia positivamente con el rendimiento ante diferentes pruebas neuropsicológicas. Los efectos de la educación se muestran en casi todas las pruebas que implican habilidades cognitivas (Lezak et al., 2012), aunque se asocia en diferente medida dependiendo del dominio cognitivo evaluado (Ardila et al., 2000). La educación puede influir tanto en el rendimiento ante un test, que una persona cognitivamente intacta pero con una educación menor puede tener menos puntuación que otra persona con deterioro cognitivo pero con mayor educación; las personas con un bajo nivel educativo pueden obtener puntuaciones clasificadas como “deterioro” si basamos las normas de un test tan solo en personas cuyo nivel educativo se aproxima al de la población general (Lezak et al., 2012).

2.1.1.1.3. Sexo

Existe controversia respecto a las diferencias en el rendimiento neuropsicológico entre hombres y mujeres. Tradicionalmente se ha asociado un mejor rendimiento a las mujeres en tareas verbales y de velocidad psicomotora, y a los hombres en tareas visuoespaciales y matemáticas (Kimura, 1999; Kimura y Clarke, 2002; Lezak et al., 2012; Munro et al., 2012; Weiss et al., 2003), aunque parece que estas últimas diferencias

podrían decrecer con una educación en inteligencia visuoespacial temprana (Reilly et al., 2017). Si bien es cierto que se han encontrado diferencias en algunos dominios cognitivos en favor de uno u otro género, estas diferencias, en ocasiones, no son tan grandes como para implicar datos normativos separados (Lezak et al., 2012). El pequeño tamaño del efecto encontrado entre las diferencias sugiere que el solapamiento entre las puntuaciones de hombres y mujeres es mayor que las diferencias entre ellas (Weiss et al., 2003); esta cuestión todavía no está clara y está lejos de estar resuelta (Lezak et al., 2012).

2.1.1.1.4. Cultura, raza y etnia

La cultura ha demostrado ser una variable con gran influencia en los resultados de diferentes pruebas de evaluación, sobre todo en aquellas pruebas cuyos estímulos o preguntas pueden estar más asociados a unas culturas que a otras, por ejemplo cuando las preguntas implican conocimiento general (ej., ¿cuál es la capital de Polonia?) o patrones de conducta (ej., comportamientos asociados a ser un buen hijo) (van de Vijver y Tanzer, 2004). Además, en algunos test, como los test de memoria verbal, se crea un gran sesgo cuando palabras de pruebas desarrolladas en una cultura se traducen directamente sin tener en cuenta que la frecuencia de uso y la utilización de una palabra puede variar ampliamente entre culturas (Lim et al., 2009). Una evaluación neuropsicológica apropiada debe tener en cuenta la cultura y la lengua de la persona a la que se evalúa, para que la evaluación clínica sea eficaz y los resultados de las investigaciones sean fiables (Cherner et al., 2007).

Asimismo, algunos autores han encontrado diferencias significativas en la puntuación de algunas pruebas neuropsicológicas entre diferentes razas/etnias (Heaton et al., 2003; Norman et al., 2011), pese a ello, son datos que se deberían tomar con prudencia, ya que en muchas ocasiones se están pasando por alto otros factores culturales y/o educativos subyacentes a los términos de raza/etnia, o incluso se están confundiendo

estos términos con la cultura, idioma o factores ambientales y/o socioeconómicos (Lezak et al., 2012; Manly y Echemendia, 2007).

2.1.1.2. Variables que influyen en las evaluaciones longitudinales sucesivas.

De la misma forma que en la teoría clásica de los test la puntuación observada como resultado de una evaluación es el resultado de la combinación entre un resultado real y un error, un cambio observado en la puntuación de evaluaciones sucesivas puede ser una combinación entre un cambio real y un error (Duff, 2012). Es importante tener en cuenta que las causas del cambio en la puntuación obtenida en evaluaciones repetidas son multifactoriales y no relacionadas con un único constructo (Heilbronner et al., 2010). Cuando se llevan a cabo evaluaciones sucesivas, se deben tener en cuenta algunos factores que pueden influir, como el efecto de la práctica, la fiabilidad de la prueba y el error de medida, y la regresión a la media (McCaffrey y Westervelt, 1995).

2.1.1.2.1. Efecto de la práctica

El efecto de la práctica se da cuando los pacientes que precisan de un seguimiento longitudinal aprenden la forma en la que se administra la prueba y los estímulos específicos a los que han sido expuestos (Benedict y Zgaljardic, 1998), y se observan mejorías en el rendimiento debidas a la experiencia previa con la prueba y no a una mejoría real (Nieto et al., 2014). Este efecto es diferente para cada prueba y/o dominio cognitivo (Rönnlund y Nilsson, 2006; Watkins y Smith, 2013), pero las pruebas de memoria están entre los tests cognitivos más influidos (Nieto et al., 2014).

En evaluación cognitiva el efecto de la práctica se ha observado hasta en lapsos de tiempo de 2 años y medio (Salthouse, 2010). Este efecto tiene una gran influencia y puede complicar la interpretación de los resultados de una administración repetida (Hawkins et al., 2004); si las pruebas se administran sin tener en cuenta este efecto, se

podría interpretar una mejoría que en realidad es inexistente o pasar por alto un deterioro que unido al efecto de la práctica no se percibe claramente.

La necesidad de administrar una prueba con el objetivo de realizar un seguimiento longitudinal a un paciente es frecuente en la práctica clínica, por ejemplo, para valorar cambios en pacientes con deterioro cognitivo, o valorar la mejoría tras un tiempo de tratamiento (Sánchez-Benavides et al., 2016). Sin embargo, debido al efecto de la práctica y sus consecuencias, no se aconseja la readministración de algunas de las pruebas neuropsicológicas disponibles (Lezak et al., 2012; Strauss et al., 2006). La disponibilidad de pruebas con versiones alternativas que nos ayuden a paliar este efecto es limitada para población española.

2.1.1.2.2. Fiabilidad y error de medida del instrumento.

La fiabilidad es una cualidad de los test, que puede definirse como el grado en el que un instrumento es consistente y está libre de error de medida; indica en qué medida una puntuación representa la puntuación verdadera (McCaffrey et al., 2000; Urbina, 2004). El error de medida se puede definir como la fluctuación en las puntuaciones resultante de factores relacionados con el proceso de medición, pero irrelevantes para lo que se está midiendo; como consecuencia de las limitaciones de los instrumentos, cuando medimos variables continuas el error de medida es inevitable (Urbina, 2004). Una parte de la diferencia en el rendimiento de un paciente entre evaluaciones puede deberse a este error (McCaffrey y Westervelt, 1995).

2.1.1.2.3. Regresión a la media.

La regresión a la media es un fenómeno estadístico que ocurre cuando una prueba es readministrada, y hace referencia a la tendencia de las puntuaciones extremas por acercarse a una puntuación más cercana a la media (Barnett, 2004; Kazdin, 1998). El incremento o decremento de las puntuaciones debido a este efecto dependerá de si la

puntuación extrema estaba por debajo o por encima de la media, respectivamente, sin necesidad de que intervenga ningún factor (McCaffrey et al., 2000; McCaffrey y Westervelt, 1995). Cuanto más extrema sea la puntuación, más probable será que se produzca una regresión a la media, sin embargo, es importante tener en cuenta cuando un cambio desafía este efecto, por ejemplo, cuando una puntuación extrema se mantiene o se vuelve aún más extrema (Duff, 2012).

2.2. Métodos de comparación de las puntuaciones del rendimiento.

La evaluación neuropsicológica de un paciente proporciona puntuaciones directas. Estas puntuaciones no se pueden interpretar por sí mismas, sino que deben ser comparadas con otras puntuaciones de referencia (Zachary y Gorsuch, 1985); para poder evaluar un déficit se establece un nivel previo o ideal de funcionamiento normal con el que poder comparar los resultados del paciente (Lezak et al., 2012). Según Lezak, la evaluación neuropsicológica, dependiendo de su propósito y de las características del paciente, utiliza dos tipos de métodos de comparación: 1) Comparación normativa, a partir de una población representativa; y 2) Comparación individual, a partir de la historia del paciente.

2.2.1. Comparación normativa.

La comparación normativa ayuda a dar sentido a las puntuaciones y a la clasificación de los evaluados, estableciendo su posición con respecto a un grupo normativo con el que comparte variables sociodemográficas como la edad, la educación y el sexo (Martínez Arias et al., 2006). Para que los resultados de una prueba sean fiables, además de estas variables sociodemográficas, se deben tener en cuenta la cultura y el lenguaje de la población para la cual las normas se han creado (Cherner et al., 2007); las

diferencias culturales ponen de manifiesto la limitación de los datos normativos a la muestra para la que han sido creados (Ryan et al., 2020).

La selección de un conjunto de datos normativos apropiados es imprescindible para la correcta evaluación neuropsicológica. Elegir unos datos normativos inadecuados es tan inconveniente como la elección de una prueba con baja fiabilidad o validez, ya que, dependiendo de los datos seleccionados para la creación de las normas, puede existir mucha variabilidad en las puntuaciones (Lezak et al., 2012; Mitrushina et al., 2005). Para poder elegir adecuadamente las normas es imprescindible conocer el procedimiento de administración de la prueba para la muestra normativa y las características del sujeto, que son específicas de la persona evaluada, como su edad, educación y sexo (Mitrushina et al., 2005). Teniendo en cuenta que las puntuaciones están directamente vinculadas a consecuencias como diagnósticos y tratamientos, los neuropsicólogos deben tener en cuenta las características específicas de las normas que utilizan (Strauss et al., 2006).

Los datos normativos deben estar actualizados, y la muestra debe estar compuesta por sujetos sanos, que deben conformar una muestra de un tamaño adecuado para ser representativo de la población (Mitrushina et al., 2005). Para la comparación normativa se han creado dos tipos de normas: las normas basadas en frecuencias y las normas basadas en regresión.

2.2.1.1. Normas basadas en frecuencias.

Las normas basadas en frecuencias permiten obtener una distribución normal de las puntuaciones. La distribución normal de las puntuaciones se obtiene mediante la conversión de las puntuaciones directas a percentiles y a puntuaciones estandarizadas (Martínez Arias et al., 2006), entre ellas, es común la utilización de puntuaciones escalares, que tienen una media de 10 y una desviación típica de 3 (Strauss et al., 2006). Estas normas deben tener en consideración la influencia de la edad, aportando para su

correcta corrección normas niveladas, la educación, y en los casos que haya demostrado influir en el rendimiento de la prueba, el sexo (Lezak et al., 2012; Martínez Arias et al., 2006; Mitrushina et al., 2005).

2.2.1.2. Normas basadas en regresión.

Las normas basadas en regresión permiten el ajuste de las normas a las características concretas del evaluado mediante una ecuación que permite insertar los datos relativos a las variables sociodemográficas del evaluado de forma continua, y se presenta como alternativa a las tablas, que implican saltos más pronunciados al dividirse por grupos de edad (Zachary y Gorsuch, 1985). Los predictores varían entre estudios, e incluyen las variables sociodemográficas (como la edad, la educación y el sexo) que hayan demostrado influir en la medida neuropsicológica evaluada (Parmenter et al., 2010), aunque hay estudios que las amplían y tienen en cuenta otras como la edad al cuadrado (Van der Elst et al., 2006) o la inteligencia premórbida (Testa et al., 2009).

2.2.2. Comparación individual

La comparación individual, a partir de evaluaciones previas del paciente, es necesaria cuando se quiere evaluar el cambio en el rendimiento cognitivo de una persona a lo largo del tiempo (Lezak et al., 2012). En la práctica clínica es habitual y está valorada la realización de un seguimiento a los pacientes a lo largo del tiempo, ya que estas evaluaciones pueden ayudar en el diagnóstico y manejo de condiciones neurológicas y psiquiátricas (Heilbronner et al., 2010); por ejemplo, conocer la tasa de deterioro cognitivo de una persona, puede ayudar a predecir la evolución de la enfermedad (Albert et al., 2007; Pakhomov et al., 2016).

Además de los seguimientos a corto plazo, tanto en el entorno clínico como de la investigación, es común realizar evaluaciones en lapsos de tiempo alrededor de los 3 meses. Por ejemplo, es habitual realizar evaluaciones neuropsicológicas antes y después

de la aplicación de un tratamiento de rehabilitación cognitiva, que puede durar entre 3 y 6 meses, con el objetivo de probar su eficacia, y un seguimiento a largo plazo, mayor a 12 meses, para valorar el mantenimiento de los cambios generados (Gómez-Gastiasoro, Peña, et al., 2019; Zucchella et al., 2013) o realizar seguimientos a largo plazo de un tratamiento o de una enfermedad crónicas, que requieren repetir una evaluación hasta 24 meses después de haber realizado la primera (Monaci y Morris, 2012).

Las evaluaciones longitudinales, sin embargo, están influidas por el efecto de la práctica, por el error de medida inherente a los test y por el efecto de regresión a la media. El efecto de la práctica puede controlarse con grupos control, cuando se quiere demostrar la eficacia de un tratamiento, mediante el uso de pruebas con formas paralelas que nos permitan evitar el aprendizaje de los estímulos, aunque son escasos los test que cuentan con ellas, o con datos que tengan en cuenta el efecto de la práctica mediante índices para evaluar el cambio fiable (Duff, 2012).

Al mismo tiempo, los instrumentos de evaluación que se utilizan pueden tener cierto grado de error en la precisión, haciendo que circunstancias debidas al azar hagan variar ligeramente la puntuación obtenida en momentos diferentes de evaluación (Iraurgi, 2009). Teniendo esto en cuenta, se ha propuesto utilizar el error estándar de medida (EEM) de un instrumento para establecer así un rango o intervalo de confianza en el que se espera que la puntuación real caiga (McCaffrey et al., 2000). El EEM se puede calcular fácilmente a partir del coeficiente de fiabilidad de la prueba (r_{xx}) y la desviación estándar (DE) de las puntuaciones de la prueba: $EEM = DE\sqrt{1 - r_{xx}}$ (Anastasi y Urbina, 1997). La utilización del EEM permite tener en cuenta el error de medida del instrumento, y controlar así las variaciones en las puntuaciones debidas al azar.

Con todo ello, se hace patente la necesidad de obtención de métodos fiables que permitan la comparación individual, teniendo en cuenta las variables que pueden estar

afectando al cambio entre evaluaciones; sobre todo cuando se evalúan cambios cognitivos (Maassen et al., 2009). Poder contar con datos que nos permitan no sólo tener en cuenta el cambio esperado, si no evaluar su magnitud, se hace especialmente relevante.

2.2.2.1. Pruebas con formas paralelas.

El efecto de la práctica ha demostrado influir significativamente en los resultados de una administración repetida; una de las técnicas utilizadas para minimizarlo o controlarlo es el uso de test con formas paralelas (McCaffrey et al., 1992; McCaffrey et al., 2000). El efecto de la práctica es mayor cuando se usan las mismas versiones de un test que cuando se usan versiones alternativas, ya que éstas suprimen o reducen significativamente este efecto (Bird et al., 2003), por ejemplo, cuando se utilizaron las mismas versiones de las pruebas de aprendizaje y memoria Test de Aprendizaje Verbal de Hopkins-Revisado (HVLTR) y Test Breve de Memoria Visuoespacial-Revisado (BVMT-R), los participantes mejoraron significativamente, pero al usar formas alternativas de estos test, sólo se observó una mejora ligera en memoria visuoespacial (Benedict y Zgaljardic, 1998).

El uso de formas paralelas nos permite evitar el efecto de la práctica que ocurre cuando los pacientes deben memorizar los mismos ítems en más de una ocasión, pero no permite evitar el efecto de la práctica que se da cuando el paciente desarrolla mejores estrategias para responder a un test por la exposición repetida al mismo procedimiento de administración (Benedict y Zgaljardic, 1998). La mejora en la segunda administración, incluso cuando se usan pruebas alternativas puede deberse, por ejemplo, a que la experiencia previa con el formato del test anula el factor sorpresa que algunos test tienen cuando se administra un ensayo de recuerdo libre, a largo plazo, dentro de una misma evaluación (como ocurre con la Figura Compleja de Rey-Osterrieth, la Figura Compleja de Taylor, HVLTR y BVMT-R) (Delaney et al., 1992; McCaffrey et al., 2000).

El incremento en el rendimiento de un paciente cuando se usan formas paralelas se llama “efecto de sofisticación del test”, y hace referencia al grado en el que una persona tiene experiencia con la prueba, que le aventaja proporcionándole práctica, e incluso reduciendo su ansiedad ante la prueba y reforzando su confianza (McCaffrey y Westervelt, 1995; Urbina, 2004).

Pese a que los test con formas paralelas permiten evitar el aprendizaje de los estímulos a los que los pacientes son expuestos, la existencia de otros factores que puedan alterar la puntuación entre evaluaciones debe ser tomada en cuenta. Dado que las pruebas con formas alternativas pueden no remediar totalmente el efecto de la práctica, y además pueden estar sesgadas por el error de medida y por el efecto de regresión a la media, se recomienda el uso métodos para evaluar el cambio fiable (Attix et al., 2009), y establecer cuando el cambio es suficientemente relevante para superar el rango atribuible al error de medida o a la variabilidad inherente al instrumento (Benedict y Brandt, 2007).

Asimismo, cabe destacar que, a pesar de las excepciones de algunas pruebas de evaluación neuropsicológica como el *Test de Aprendizaje Verbal España-Complutense* (TAVEC) (Benedet y Alejandre, 1998; Nieto et al., 2014), el *Test de Recuerdo Verbal Selectivo* (Campo et al., 2000), o el *Free and Cued Selective Reminding Test* (Grau-Guinea et al., 2018), las pruebas neuropsicológicas para población española que cuentan con formas alternativas son todavía escasas.

2.2.2.2. Índices de cambio fiable.

La Academia Americana de Neuropsicología Clínica (American Academy of Clinical Neuropsychology; AACN) publicó un informe en 2010 en el que subrayaba que los neuropsicólogos deberían conocer el concepto y los métodos estadísticos relacionados con los índices de cambio fiables para la medición de los cambios asociados a las evaluaciones repetidas (Heilbronner et al., 2010).

Se han propuesto diferentes métodos estadísticos para valorar el cambio obtenido entre evaluaciones y si éste es o no significativo (Duff, 2012; Sánchez-Benavides et al., 2016). Entre estos métodos se encuentran diferentes índices de cambio fiable como los datos normativos relativos a las puntuaciones de discrepancia, el Índice de Desviación Estándar (IDE), el Índice de Cambio Fiable (ICF) (Jacobson y Truax, 1991) y una versión del ICF que tiene en cuenta, además, el efecto de la práctica (Chelune et al., 1993), o las fórmulas basadas en regresión (McSweeney et al., 1993).

Los índices de cambio fiables son, esencialmente, procedimientos estadísticos que tratan de tener en cuenta el efecto de la práctica, el error de medición y la regresión a la media, describiendo la distribución de las puntuaciones de cambio esperables cuando no existe un cambio real (Attix et al., 2009).

Existen varios índices de cambio fiable; estos índices, ampliamente utilizados en países de habla inglesa, no están apenas disponibles para población española, para la cual tan sólo existe una publicación (Sánchez-Benavides et al., 2016). A continuación, se presentan cinco índices de cambio fiable, que constituyen una herramienta imprescindible en la práctica clínica y de investigación.

2.2.2.2.1. Puntuaciones de discrepancia.

La puntuación de discrepancia de un paciente se calcula mediante la diferencia entre la puntuación del tiempo 2 y la del tiempo 1. La puntuación resultante se debe comparar con las de la muestra normativa, que proporcionan información sobre la frecuencia y distribución de estas puntuaciones en la población, lo que nos indica el percentil correspondiente a la puntuación de discrepancia obtenida por el paciente.

2.2.2.2.2. Índice de Desviación Estándar (IDE).

La aplicación del IDE implica el cálculo de la puntuación de discrepancia de un paciente, mediante la resta entre la puntuación obtenida en el tiempo 2 y la puntuación

obtenida en el tiempo 1, y su división entre la DE de la muestra normativa (DE_{norm}) en el tiempo 1.

$$IDE = \frac{\text{Tiempo2} - \text{Tiempo1}}{DE_{norm}}$$

2.2.2.2.3. Índice de Cambio Fiable (ICF).

El ICF es un índice estadístico propuesto por Jacobson y Truax en el año 1991 (Jacobson y Truax, 1991). Este índice nos permite comparar entre puntuaciones obtenidas por una determinada persona en dos momentos temporales diferentes, y nos proporciona información sobre si un cambio se debe o no al azar, utilizando la precisión de medida del instrumento y la variabilidad de las puntuaciones de la muestra.

La aplicación del ICF implica el cálculo de la puntuación de discrepancia de un paciente restando la puntuación en el tiempo 1 a la obtenida en el tiempo 2, y dividiendo el resultado entre el error estándar de las diferencias entre medidas (EED). El EED se obtiene a partir del EEM, y éste, a su vez, a partir de la DE_{norm} y del coeficiente de fiabilidad de la prueba (r_{xx}). La fórmula de Jacobson y Truax es la siguiente:

$$ICF = \frac{\text{Tiempo2} - \text{Tiempo1}}{EED} = \frac{\text{Tiempo2} - \text{Tiempo1}}{EEM\sqrt{2}} = \frac{\text{Tiempo2} - \text{Tiempo1}}{\sqrt{2(DE_{norm}\sqrt{1-r_{xx}})^2}}$$

$$EED = \sqrt{2(EEM)^2} = EEM\sqrt{2}$$

$$EEM = DE_{norm}\sqrt{1-r_{xx}}$$

2.2.2.2.4. Índice de Cambio Fiable + efecto de la práctica (ICF+EP).

El ICF fue diseñado para medidas de constructos psicológicos como depresión y ansiedad, pero no tiene en cuenta el efecto de la práctica que tienen las pruebas que miden el rendimiento cognitivo (Duff, 2012). Para controlar el efecto de la práctica (EP), se ha propuesto una modificación de la fórmula del RCI que tiene en cuenta este efecto en diversas pruebas de evaluación neuropsicológica (Chelune et al., 1993).

Utilizando este nuevo índice, diversos autores han proporcionado normas de corrección para evaluaciones que se han realizado con lapsos de tiempo de entre una semana y dos años entre ellas (Duff, 2014; Levine et al., 2004; Sánchez-Benavides et al., 2016).

$$ICF = \frac{\text{Tiempo2} - \text{Tiempo1} - EP}{EED}$$

2.2.2.2.5. Fórmulas basadas en regresión estandarizadas.

El método basado en regresión (McSweeny et al., 1993) permite predecir la puntuación de la segunda administración teniendo en cuenta la puntuación de la primera (método de regresión simple), y en ocasiones, otras variables sociodemográficas (método de regresión complejo), y obtener con ello una puntuación estandarizada.

2.2.2.2.6. Ventajas y desventajas de los diferentes índices de cambio fiable.

El método basado en la distribución por frecuencias de las puntuaciones de discrepancia, es el método más sencillo e intuitivo para uso en la práctica clínica, aunque es menos preciso que otros métodos por ofrecer un amplio rango de valores (Duff, 2012). Los otros cuatro métodos, en cambio permiten proporcionar puntuaciones z específicas.

Una de las ventajas del Índice de Desviación Estándar (IDE) es que es fácil de calcular, por ello es uno de los métodos más ampliamente utilizados en la práctica clínica, sin embargo, este método no permite tener en cuenta ni el efecto de la práctica ni el error de medida del instrumento.

Los índices Índice de Cambio Fiable (ICF) e Índice de Cambio Fiable+Efecto de la Práctica (ICF+PE) permiten tener en cuenta el error de medida del instrumento, aumentando así su fiabilidad. Una ventaja añadida del método ICF+EP es que permite contar con el efecto de la práctica que tienen las pruebas de evaluación neuropsicológica,

algo especialmente relevante en determinadas pruebas, que cuentan con una amplia influencia de este efecto.

Los cuatro métodos anteriores, no obstante, no permiten tener en cuenta variables sociodemográficas. Precisamente una de las grandes ventajas del método de regresión complejo es que, además de tener en cuenta el efecto de la práctica, permite incluir variables sociodemográficas como la edad, la educación y el sexo, que han demostrado influir en menor cantidad en los análisis, aunque de forma significativa (Attix et al., 2009; Duff, 2012; Sánchez-Benavides et al., 2016).

A pesar de que no existe consenso sobre cuál de los métodos es el mejor, y dependiendo de las circunstancias del paciente y de la evaluación se podría elegir uno u otro, los métodos que mejor han demostrado identificar el cambio, mostrando una mayor concordancia entre ellos, son el ICF+EP y los métodos basados en regresión (Frerichs y Tuokko, 2005; Heaton et al., 2001; Temkin et al., 1999). Esto puede deberse a que estos métodos permiten tener en cuenta el efecto de la práctica (Maassen et al., 2009).

2.2.2.2.7. Interpretación de las puntuaciones z.

El resultado de las fórmulas anteriores da lugar a una puntuación z, que permite su comparación con una tabla de distribución normal; el punto de corte para establecer cuando un cambio es clínicamente relevante difiere ligeramente de unos autores a otros.

Algunos autores establecen el punto de corte para considerar el cambio fiable en una cifra menor a -1,645 o mayor a 1,645, ya que este número corresponde a los percentiles 5 y 95 respectivamente, e implica que el 90% de las puntuaciones en una distribución normal se situarán en el rango comprendido entre -1,645 y +1,645 (Duff, 2012; Sachs et al., 2012; Temkin et al., 1999) y sólo un 5% de las personas sanas caerán por azar fuera de este rango en cada dirección (Heaton et al., 2001).

Para Jacobson y Truax, si el resultado de la operación da lugar a una cifra menor a -1,96 o mayor a 1,96 (número que equivale a dos desviaciones típicas en las puntuaciones z) se puede decir que el cambio entre la primera y la segunda evaluación es fiable y real. Otros autores argumentan que es responsabilidad del clínico establecer el punto de corte más adecuado (Sánchez-Benavides et al., 2016).

2.3. Normalización en población española

La necesidad de obtención de datos adecuados a las características sociodemográficas y culturales ha quedado sobradamente demostrada. En nuestro caso, además, está ampliamente reconocida la necesidad de obtener datos normativos adecuados a la población española para pruebas de evaluación neuropsicológica (Jordi Peña-Casanova, Blesa, et al., 2009). Como explican Del Pino, Peña, Schretlen, et al. (2015), muchos de los datos normativos empleados en la evaluación neuropsicológica en España no han sido adaptados a esta población, sino que se han normalizado en población de habla inglesa o hispanohablante procedente de Estados Unidos y Sudamérica, lo que implica una considerable disminución en su representatividad. Además, muchas de las normas adaptadas a las características de la sociedad española, están restringidas a grupos de edad y características específicas y limitadas (del Pino, Peña, Schretlen, et al., 2015; Mitrushina et al., 2005).

Ante la necesidad de normas adaptadas a población española, en 2009 se creó el proyecto Neuronorma, con el objetivo de obtener datos psicométricos y normativos para población española mayor de 49 años (Jordi Peña-Casanova, Blesa, et al., 2009) y posteriormente, el proyecto Neuronorma Jóvenes, que incluye también a personas de 18 a 49 años (Jordi Peña-Casanova et al., 2012). Asimismo, y teniendo en cuenta la necesidad de obtención de datos que permitan evaluar e interpretar de forma fiable el

cambio cognitivo observado entre evaluaciones, en el marco del proyecto Neuronorma, en 2016, se publicó un estudio que permite tener en cuenta el cambio esperado tras un lapso de tiempo de un año entre dos evaluaciones sucesivas para personas de 50 a 85 años de edad (Sánchez-Benavides et al., 2016).

La contribución de este proyecto a la obtención de datos normativos e índices de cambio fiable para población española ha sido muy relevante, y ha favorecido la cobertura parcial de la necesidad existente, sin embargo, muchas pruebas ampliamente utilizadas no están incluidas. Entre estas pruebas de uso habitual se pueden incluir por ejemplo el Test de Clasificación de Tarjetas de Wisconsin - Modificado (M-WCST) (Schretlen, 2010), la Figura Compleja de Taylor (FCT) (Taylor, 1969), el HVLt-R (Benedict et al., 1998; Brandt y Benedict, 2001) o el BVMT-R (Benedict, 1997; Benedict et al., 1996).

2.4. Normacog

El proyecto Normacog tiene como objetivo tanto la normalización y baremación de pruebas de evaluación neuropsicológica adaptadas a las características sociodemográficas de la población española actual (del Pino, Peña, Schretlen, et al., 2015) como la obtención de índices de cambio fiable para estas pruebas. Este proyecto se puso en marcha en 2012 para dar respuesta a la carencia existente de pruebas adaptadas para la población española, iniciándose como una colaboración entre el equipo Neuropsicología de los Trastornos Médicos Severos (I.P. Dra. Natalia Ojeda) y Johns Hopkins Medical Institutions – Division of Medical Psychology- (I.P. Dr. David Schretlen).

2.4.1. Protocolo Normacog

La batería de pruebas incluida en este proyecto es una batería de evaluación neuropsicológica que consta de un total de 19 pruebas ampliamente conocidas y utilizadas en evaluación neuropsicológica tanto nacional como internacionalmente (ver Tabla 1). Estas pruebas fueron seleccionadas teniendo en cuenta las pruebas incluidas en el Sistema Normativo de Calibrado Neuropsicológico [CNNS; Calibrated Neuropsychological Normative System (Schretlen et al., 2010)] y pruebas de uso extendido para las cuales no existían baremos normativos en nuestra población. Además, se incluyó una nueva versión de Test de Stroop, el test UD Interferencia (Ojeda et al., 2014), con el objetivo de evitar algunas de las limitaciones existentes en otras versiones relacionadas con el daltonismo y las dificultades de lectura en la tercera edad.

Estas pruebas permiten una exhaustiva evaluación ya que obtienen datos relativos a diferentes procesos cognitivos (atención, memoria visuoespacial y verbal, funciones ejecutivas, fluidez verbal y lenguaje, concentración, capacidades visuoconstructivas, orientación, velocidad de procesamiento, etc.).

La baremación y normalización de este protocolo se tomó como punto de partida de otro objetivo más ambicioso: el cálculo del efecto del aprendizaje en evaluaciones repetidas y la aportación de herramientas como los índices de cambio fiable.

Tabla 1.*Batería de Pruebas Normacog (del Pino, Peña, Schretlen, et al., 2015).*

Test	Autor (año)	Áreas de evaluación
Prospective Memory Test	Einstein y McDaniel (1990)	Memoria prospectiva
Test de acentuación de palabras	Del Ser et al (1997)	Rendimiento premórbido
Subprueba de pseudopalabras del PROLEC	Cuetos et al (2007)	Procesos lectores y fonológicos
Evaluación cognitiva de Montreal	Nasreddline et al (2003)	Disfunciones cognitivas leves
Escala de depresión geriátrica	Yesavage et al (1982)	Depresión
Figura compleja de Taylor	Taylor (1969)	Capacidad visuoespacial y memoria visual
Test UD interferencia	Ojeda et al (2013)	Atención y resistencia a la interferencia
Test de comparación perceptual	Salthouse (1991)	Velocidad de percepción
Test de Aprendizaje Verbal de Hopkins-Revisado. Versiones 2 y 4	Brandt y Benedict (2001)	Aprendizaje y memoria verbal
Test Breve de Memoria Visuoespacial-Revisado. Versiones 1 y 3	Benedict (1997)	Aprendizaje y memoria visual
Test modificado de clasificación de tarjetas de Wisconsin	Schretlen (2010)	Funciones ejecutivas y flexibilidad mental
Trail Making Test	Reitan (1985)	Atención dividida, flexibilidad y función ejecutiva
Grooved Pegboard	Test Klove (1963)	Destreza manipulativa y velocidad de procesamiento
Test de estimación cognitiva	Shallice y Evans (1978)	Función ejecutiva
Test del dibujo del reloj	Goodglass y Kaplan (1972)	Capacidades visuoespaciales y ejecutivas
Test breve de atención	Schretlen et al (1996)	Atención dividida y memoria de trabajo
Subtest de fluidez verbal. Calibrated Ideational Fluency Assessment	Schretlen y Vannorsdall (2010)	Fluidez verbal
Test de denominación de Boston	Goodglas y Kaplan (1986)	Denominación verbal
Cuestionario de actividades de la vida diaria de Lawton	Lawton y Brody (1969)	Habilidades de la vida diaria

2.4.1.1. Datos normativos disponibles

Desde 2012 y hasta el inicio de esta tesis se han publicado baremos normativos adaptados a la población española para seis pruebas diferentes, con el objetivo de ponerlos a disposición de la comunidad científica y profesional y aumentar así el número y la calidad de los datos normativos disponibles para nuestra población. Se han creado datos normativos para las siguientes pruebas:

- Test de Acentuación de Palabras (TAP) y Subprueba de lectura de pseudopalabras del PROLEC-R (Cuetos et al., 2014; del Pino et al., 2018; Del Ser et al., 1997).
- Evaluación cognitiva de Montreal (MoCA) (Nasreddine et al., 2005; Ojeda, del Pino, et al., 2016).
- Figura compleja de Taylor (FCT) (del Pino, Peña, Ibarretxe-Bilbao, et al., 2015; Taylor, 1969).
- Test de Clasificación de Tarjetas de Wisconsin - Modificado (M-WCST) (del Pino et al., 2016; Schretlen, 2010).
- Test de Comparación Perceptual de Salthouse (TCPS) (Peña et al., 2016; Salthouse y Babcock, 1991).

2.4.1.2. HVLТ-R

Una parte imprescindible de la evaluación neuropsicológica ha sido siempre la evaluación del aprendizaje y la memoria verbal, por ser dominios afectados en muchas patologías (Benedict et al., 1998; Norman et al., 2011).

A la hora de seleccionar un test de evaluación, se debe tener en cuenta, por un lado, el efecto del aprendizaje que tienen las pruebas de memoria., y por otro, que las características sociodemográficas de la población y el idioma de administración de los

test influyen significativamente en los resultados de las pruebas (Cherner et al., 2007; Lim et al., 2009; Nieto et al., 2014). Por ello, es importante contar con test fiables que se adapten a las circunstancias de la evaluación y que se adecúen a las características de la población en la que es administrado. Pese a ello, son escasas las pruebas que permiten una reevaluación breve del aprendizaje y la memoria verbal disponibles para población española.

Entre los instrumentos comúnmente utilizados a nivel internacional para la evaluación del aprendizaje y la memoria verbal se encuentra el HVLTR (Brandt y Benedict, 2001), publicada precisamente con el objetivo de proporcionar un test breve y fiable de aprendizaje y memoria verbal con versiones paralelas que permitieran su readministración en seguimientos longitudinales. HVLTR cuenta con un total de 6 versiones alternativas; los autores recomiendan que en seguimientos longitudinales se empleen conjuntamente las versiones 1, 2 y 4 o las versiones 3, 5 y 6 (Benedict et al., 1998). Este test presenta una excelente sensibilidad (87–96%) y especificidad (80-98%) a la hora de discriminar entre población sana y con demencia (Frank y Byrne, 2000; Hogervorst et al., 2002) y cuenta con muy buena validez concurrente y de constructo (Shapiro et al., 1999).

La prueba HVLTR presenta varias ventajas, como evitar el cansancio por parte del paciente o disminuir el tiempo de trabajo del examinador, ambas especialmente interesantes cuando la prueba se administra en el contexto de una batería más amplia, o cuando trabajamos con personas mayores donde la fatiga influye sobre el rendimiento (Lacritz et al., 2001).

Características sociodemográficas como la edad y la educación han demostrado influir en el rendimiento en el HVLTR (Arango-Lasprilla et al., 2015; Benedict et al., 1998; Friedman et al., 2002; Hester et al., 2004; Miotto et al., 2012; Norman et al., 2011;

Ryan et al., 2020; Waldrop-Valverde et al., 2015). Además, se ha comprobado que las mujeres obtienen una puntuación superior en algunas de las variables del test (Friedman et al., 2002; Norman et al., 2011; Ryan et al., 2020; Vanderploeg et al., 2000), ventaja que desaparece cuando se comparan grupos de personas con déficits cognitivos (Brunet et al., 2019).

Las características y ventajas que ofrece este test han contribuido a que la prueba se haya traducido a diferentes idiomas y se hayan obtenido datos normativos en diferentes países (Arango-Lasprilla et al., 2015; Benedict et al., 1998; Cherner et al., 2007; Friedman et al., 2002; Hester et al., 2004; Miotto et al., 2012; Norman et al., 2011; Ryan et al., 2020; Vanderploeg et al., 2000; Waldrop-Valverde et al., 2015). Además, HVLT-R se ha utilizado para evaluar a pacientes con diferentes patologías neuropsiquiátricas (Gómez-Gastiasoro, Zubiaurre-Elorza, et al., 2019; Schretlen et al., 2007), demencia (Frank y Byrne, 2000; Gonzalez-Palau et al., 2013; Hogervorst et al., 2002) y daño cerebral adquirido (Neil-Pirozzi et al., 2012).

Este test, por sus destacadas características, ha sido seleccionado para formar parte de diferentes baterías internacionales de evaluación neuropsicológica, como la batería *Calibrated Neuropsychological Normative System* (CNNS) (Schretlen et al., 2010), y fue la prueba de selección para la medición del aprendizaje y memoria en la batería *Measurement and Treatment Research to Improve Cognition in Schizophrenia* (MATRICS) (Nuechterlein y Green, 2006).

La amplia utilización del HVLT-R en evaluación neuropsicológica pone de manifiesto la importancia de la obtención de datos normativos adecuados. Sin embargo, hasta la fecha, no contamos con datos normativos adaptados a población española.

2.4.1.3. BVMT-R

La evaluación de la memoria visuoespacial es especialmente relevante en una evaluación neuropsicológica, ya que esta es una de las funciones cognitivas más afectadas en muchas patologías que cursan con deterioro cognitivo y también en envejecimiento normal (Duff, 2016; Kane y Yochim, 2014).

Existen ciertas características que se deben tener en cuenta a la hora de seleccionar un test de evaluación del aprendizaje y la memoria visual. Además del efecto de la práctica, que influye en las evaluaciones sucesivas, se debe tener en cuenta que la medición del aprendizaje requiere suficientes ensayos para permitir la aparición de una curva de aprendizaje (Duff et al., 2011; Lezak et al., 2012; Nieto et al., 2014). Sin embargo, son escasas las pruebas para población española que cuentan con diferentes ensayos de aprendizaje, con versiones paralelas y que permiten una reevaluación breve de las capacidades de aprendizaje y memoria.

El test BVMT-R (Benedict, 1997) es uno de los instrumentos comúnmente utilizados para la evaluación del aprendizaje y la memoria visuoespacial. Esta prueba se creó teniendo en cuenta el efecto de la práctica observado en pacientes en rehabilitación, y con el objetivo dar respuesta a la carencia de pruebas con versiones alternativas que permitieran su readministración (Benedict, 1988). Este test presenta una excelente sensibilidad (98%) y especificidad (82%) a la hora de identificar pacientes sanos y con demencia (Benedict et al., 1996) y una excelente validez de constructo (Benedict, 1997). Las seis versiones alternativas incluidas en BVMT-R han demostrado reducir el efecto del aprendizaje tanto en participantes sanos (Benedict y Zgaljardic, 1998; Zgaljardic y Benedict, 2001) como en pacientes con patología (Benedict, 2005).

Este test constituye una herramienta breve que permite disminuir el tiempo de administración, convirtiéndolo en una herramienta ideal cuando se utiliza en contextos

clínicos con pacientes que presentan fatiga o para ser incluida como parte de una batería más amplia. La brevedad del test, la disponibilidad de versiones paralelas y la inclusión de tareas de aprendizaje y reconocimiento además del recuerdo a largo plazo, son grandes ventajas de este test (Lezak et al., 2012).

El rendimiento cognitivo en el BVMT-R parece estar influido por características sociodemográficas, como la edad y la educación (Benedict et al., 1996; Duff, 2016; Kane y Yochim, 2014; Kemmotsu et al., 2013; Miotto et al., 2012), pero no por el sexo (Benedict et al., 1996; Brunet et al., 2019; Duff, 2016; Kane y Yochim, 2014; Kemmotsu et al., 2013; Miotto et al., 2012) aunque en algunos estudios las mujeres puntuaron ligeramente más alto en algunas variables (Gale et al., 2007, 2016).

Este test se ha validado con población sana en diferentes países, ajustándose así los datos normativos a la población de cada país (Argento et al., 2016; Cherner et al., 2007; Duff, 2016; Gale et al., 2007; Kane y Yochim, 2014; Miotto et al., 2012; Norman et al., 2011). Además, ha sido utilizado para evaluar a personas con diferentes patologías neurodegenerativas como enfermedad de Parkinson (Lucas-Jiménez et al., 2015) o esclerosis múltiple (Argento et al., 2016), trastornos psiquiátricos como esquizofrenia o trastorno bipolar (Schretlen et al., 2007), demencias y deterioro cognitivo (Wasserman et al., 2019), y envejecimiento sano (Brunet et al., 2019; Montoya-Murillo et al., 2019).

El BVMT-R, por sus propiedades destacadas, ha sido incluido en diferentes baterías de consenso internacional de evaluación neuropsicológica, como la batería para la evaluación de la cognición en esquizofrenia *Mental Healths' Measurement and Treatment Research to Improve Cognition in Schizophrenia* (MATRICS) (Nuechterlein et al., 2008; Nuechterlein y Green, 2006), en las baterías para la evaluación de los déficits cognitivos en la esclerosis múltiple *Minimal Assessment of Cognitive Function In Multiple Sclerosis* (MACFIMS) (Benedict et al., 2006; Parmenter et al., 2010) y *Brief*

International Cognitive Assessment for Multiple Sclerosis (BICAMS) (Langdon et al., 2012) y en la batería de evaluación neuropsicológica *Calibrated Neuropsychological Normative System (CNNS)* (Schretlen et al., 2010).

BVMT-R es un test ampliamente utilizado en evaluación neuropsicológica, por ello, resulta fundamental contar con datos normativos adaptados a las características sociodemográficas de la población a la que se evalúa. Sin embargo, hasta el momento, este test no cuenta con datos normativos adaptados a población española.

2.4.1.4. UD Interferencia

El test de Stroop es un instrumento ampliamente utilizado a nivel mundial tanto en la práctica clínica como en investigación (Lezak et al., 2012; Mitrushina et al., 2005). Este test evalúa la velocidad de procesamiento, la atención selectiva, la inhibición y la flexibilidad cognitiva, y la resistencia a la interferencia (Jordi Peña-Casanova, Quiñones-Úbeda, et al., 2009; Rivera et al., 2017; Sisco et al., 2016; Strauss et al., 2006).

El origen de este test se basa en el estudio llevado a cabo por Stroop en 1935 (Stroop, 1935), que ha sido de referencia en el campo de la psicología cognitiva y la neuropsicología desde entonces. El paradigma de Stroop implica la comparación entre la ejecución de tareas sin distractores y la ejecución de una tarea en presencia de un distractor; siguiendo este paradigma, se han creado diferentes versiones del test de Stroop (Lezak et al., 2012; Strauss et al., 2006) y aunque no hay una versión reconocida como standard, hay varias versiones comercialmente disponibles o fácilmente accesibles (Mitrushina et al., 2005).

Estas versiones difieren entre ellas en varios aspectos, por ejemplo, el número de ítems incluidos en cada lámina, las formas utilizadas en la lámina de color, o el número y orden de administración de las láminas. Una de las características que comparten, es que todas incluyen entre sus colores el rojo, azul y verde. Esto obstaculiza la evaluación

para personas con dificultades a la hora de discriminar colores, como las personas mayores, que tienen dificultades para discriminar entre los colores azul y verde (Pachana et al., 2004), e impide la utilización de la prueba con personas con daltonismo que tienen dificultades para discriminar entre los colores rojo y verde (Rivera et al., 2017; Strauss et al., 2006). Además, el elevado número de ítems en cada lámina que incluyen algunas versiones, y su pequeña tipografía, puede complicar la tarea por la dificultad añadida a la hora de distinguir los ítems con claridad, lo que afectaría al rendimiento sobre todo en personas mayores o con problemas en la visión (Pachana et al., 2004).

Teniendo en cuenta el extenso uso de este test y algunas de las limitaciones que presentan las actuales versiones, en el marco del proyecto Normacog se ha creado una nueva versión del test, denominado UD Interferencia (Ojeda et al., 2014), utilizando colores fácilmente distinguibles (azul, negro y rosa), un tamaño de letra más grande del utilizado en otras versiones y un reducido tiempo de administración.

En conclusión, la influencia de las variables sociodemográficas en el rendimiento de las pruebas de evaluación neuropsicológica es evidente, por ello, resulta fundamental que los baremos que se utilizan para la corrección se obtengan de la población sobre la que se empleará la prueba, para evitar sesgos y errores de interpretación. Sin embargo, no contamos con datos normativos adaptados a población española para algunas pruebas de referencia. Además, y comprobada la necesidad de readministración de pruebas de evaluación neuropsicológica, resulta fundamental la obtención de datos que nos permitan corregir adecuadamente estas evaluaciones sucesivas, evitando sesgos asociados a la interpretación del cambio entre evaluaciones, mediante datos que nos proporcionen, además, información sobre la magnitud del cambio observado.

III. Acercamiento al presente estudio, objetivos e hipótesis

3. Acercamiento al presente estudio, objetivos e hipótesis.

La presente tesis incluye cuatro estudios que pretenden mejorar la disponibilidad de datos normativos para población española adulta, y además, proporcionar índices de cambio fiable para la correcta interpretación del cambio entre evaluaciones sucesivas.

3.1. Estudio I

“Test de Aprendizaje Verbal de Hopkins-Revisado: normalización y estandarización de la prueba en población española”

Contexto

La prueba de aprendizaje y memoria verbal HVLT-R no cuenta hasta el momento con datos normativos adecuados a la población española que permitan corregir adecuadamente esta prueba. Considerando que HVLT-R es una prueba ampliamente utilizada tanto a nivel nacional como internacional, se hace especialmente relevante la obtención de datos normativos adecuados a nuestra población.

Objetivos

- El objetivo principal de este estudio fue explorar la relación entre HVLT-R y las variables sociodemográficas edad, años de educación y sexo, y proporcionar datos normativos y estandarizados para el test HVLT-R, adaptado a las características sociodemográficas de la población española actual.

Hipótesis

- Las variables sociodemográficas edad y años de educación mostrarán un efecto significativo en el rendimiento ante HVLT-R, por lo que se ofrecerán baremos que permitan tener en cuenta estas características. El sexo tendrá tan sólo una ligera influencia en el rendimiento ante el test.

3.2. Estudio II

“Test Breve de Memoria Visuoespacial-Revisado: normalización y estandarización de la prueba en población española”

Contexto

La prueba de aprendizaje y memoria visuoespacial BVMT-R no cuenta hasta el momento con datos normativos adecuados a nuestra población. Teniendo en cuenta que BVMT-R es una prueba ampliamente utilizada tanto de forma individual como en el marco de diferentes baterías de evaluación neuropsicológica, a nivel nacional e internacional, es destacable la necesidad de obtención de datos normativos para población española.

Objetivos

- El objetivo principal de este estudio fue explorar la relación entre BVMT-R y las variables sociodemográficas edad, años de educación y sexo, y proporcionar datos normativos y estandarizados para el test BVMT-R, adecuados a las características sociodemográficas de la población española.

Hipótesis

- Las variables sociodemográficas edad y años de educación mostrarán un efecto significativo en el rendimiento ante BVMT-R, por lo que los baremos normativos permitirán tener en cuenta estas características. El sexo no tendrá influencia en el rendimiento ante este test.

3.3. Estudio III

“Test UD Interferencia: creación y validación de un nuevo instrumento de resistencia a la interferencia. Normalización y estandarización en población española”

Contexto

Las actuales versiones del Test de Stroop tienen algunas limitaciones cuando se administran a personas con daltonismo o a personas mayores con dificultades de visión. Se identifica la necesidad de un nuevo instrumento que, teniendo en cuenta estas limitaciones, ofrezca alternativas a estas situaciones clínicas, y la obtención de datos normativos y estandarizados para éste adaptados a población española.

Objetivos

- El objetivo principal de este estudio fue presentar una versión alternativa del test de Stroop, el UD Interferencia, que presenta algunas ventajas sobre anteriores versiones, y ofrecer sus propiedades psicométricas en cuanto a fiabilidad y validez.
- El segundo objetivo fue proporcionar datos normativos y estandarizados para el test UD Interferencia, adecuando las normas a la población española adulta de 18 a 93 años de edad.

Hipótesis

- El test UD Interferencia mostrará una buena fiabilidad, mediante una alta consistencia interna y una buena correlación test-retest.
- El test UD Interferencia mostrará una buena validez concurrente al evaluar su relación con otras medidas de velocidad de procesamiento, atención, interferencia y/o flexibilidad.
- El test UD Interferencia mostrará una buena validez de constructo al evaluar su relación con los constructos atención, velocidad psicomotora, fluidez verbal y aprendizaje y memoria verbal.

- Las variables sociodemográficas edad y años de educación mostrarán un efecto significativo en el rendimiento ante UD Interferencia, por ello, los baremos normativos permitirán tener en cuenta estas características. El sexo tendrá tan sólo una ligera influencia en alguna variable del test.

3.4. Estudio IV

“Índices de cambio fiable para la evaluación del cambio en 16 pruebas de evaluación neuropsicológica en seis momentos temporales diferentes”

Contexto

En evaluación neuropsicológica es frecuente la necesidad de administrar una prueba en repetidas ocasiones, pero la interpretación del cambio entre evaluaciones puede verse sesgada por el efecto de la práctica, el error de medida y/o por el efecto de regresión a la media. Los índices de cambio fiable pueden ayudar a paliar estos sesgos, pero son escasos los estudios que proporcionan estos índices a nivel internacional, aún más, para población española.

Objetivos

- El objetivo principal de este estudio es proporcionar datos que permitan una adecuada interpretación de los cambios obtenidos entre evaluaciones, adaptados a la población española actual, para 16 pruebas de uso frecuente en neuropsicología en seis momentos diferentes de evaluación sucesiva, mediante los cinco índices de cambio fiable descritos con anterioridad en este trabajo.

Hipótesis

- Las diferencias entre la primera y la segunda evaluación serán mayores cuando la evaluación se ha repetido con un menor lapso de tiempo, y se diluirán en el grupo evaluado tras 24 meses.

- La puntuación de los test será mejor en la segunda evaluación, debido al efecto de la práctica.
- La obtención de índices de cambio fiable nos permitirá ofrecer una herramienta necesaria para la correcta interpretación de los cambios entre evaluaciones neuropsicológicas.

IV. Métodos

4. Métodos

La presente tesis se compone de cuatro estudios que proporcionan datos para la correcta corrección de diferentes pruebas neuropsicológicas. Todos los estudios cumplen las directrices internacionales en materia de investigación biomédica con seres humanos (Declaración de Helsinki, Edimburgo 2000) y los datos obtenidos se han tratado siguiendo la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal.

Estos estudios están compuestos por diferentes muestras y los análisis estadísticos que incluyen son diferentes en cada uno de ellos. Una descripción detallada de las características de las muestras, los test utilizados, los aspectos metodológicos, y el análisis de datos se encuentra a continuación.

4.1. Muestra de los estudios

La muestra de este proyecto multicéntrico se compone de personas sanas de diferentes comunidades autónomas del país (Galicia, Cantabria, País Vasco, Comunidad Valenciana, Murcia, Castilla-La Mancha, Andalucía, Castilla y León y Madrid), con el fin de intentar la representación de la población española.

Teniendo en cuenta el tamaño de la población española mayor de 18 años según los datos publicados por el Instituto Nacional de Estadística (INE; enero de 2019), se calculó con el programa Epi Info (Dean et al., 1996) el tamaño muestral necesario para obtener una muestra representativa (mínimo necesario: 500 personas) de la población española. Se decidió aumentar ese número al máximo posible para aumentar su representatividad en los estudios de normalización, contando así con 700 participantes para el estudio I, y aumentando este número hasta 905 para los estudios II y III.

Los criterios de inclusión y exclusión de los participantes en estos estudios fueron los siguientes: a) *Criterios de inclusión*: personas de ambos sexos entre 18 y 90 años; lengua materna castellano o bilingüe; sujetos representativos según los datos publicados por el INE; participación voluntaria y firma del consentimiento informado; capacidad suficiente de lectoescritura; y ausencia de deterioro cognitivo, evaluado mediante la Entrevista Telefónica del Estado Cognitivo (TICS) (Brandt et al., 1988); b) *Criterios de exclusión*: historial previo de enfermedad física o psíquica que comprometiera significativamente el funcionamiento cognitivo de la persona, evaluado mediante una entrevista clínica estructurada; limitación sensorial (visual o auditiva) que no pudiera ser compensada con correctivos (gafas o audífonos); historia significativa de abuso de alcohol u otros tóxicos; falta de voluntad o incapacidad de la persona para colaborar en el estudio; y no completar las pruebas o limitada información sobre los datos necesarios para la selección de los participantes.

4.1.1 Estudio I.

La muestra incluida en el estudio I estaba formada por 700 participantes sanos (305 hombres y 395 mujeres), de entre 18 y 90 años, con rangos de educación de 0 a 25 años. Dentro de esta muestra, se seleccionó un subgrupo de 42 personas que completaron también una evaluación longitudinal (media de $1,96 \pm 1,53$ meses) que incluía una versión paralela de HVLT-R (la 2 cuando en la primera evaluación se había administrado la 4 y viceversa).

4.1.2 Estudios II y III.

La muestra incluida en los estudios II y III estaba formada por 905 participantes sanos (401 hombres y 504 mujeres) de edades comprendidas entre 18 y 93 años, con rangos de educación de 0 a 25 años. De los 905 participantes, 2 no fueron evaluados con BVMT-R y otros 23 no completaron la prueba de reconocimiento y 25 no fueron

evaluados con UD Interferencia. Para el estudio III, se seleccionaron dos subgrupos de personas de entre las 905; 188 personas completaron también una evaluación longitudinal entre 1 y 6 meses (media de $3,45 \pm 2,16$ meses) y 185 personas la completaron entre 9 y 24 meses (media de $16,05 \pm 7,05$ meses).

4.1.3 Estudio IV.

La muestra incluida en el estudio IV se obtuvo a partir de la evaluación longitudinal de un subgrupo de participantes del proyecto multicéntrico Normacog; partiendo de una muestra original de 905 participantes, se readministró la batería de pruebas Normacog a 388 personas sanas (162 hombres y 226 mujeres), de entre 18 y 84 años, con rangos de educación de 0 a 25 años. Los participantes se dividieron en seis grupos de evaluación longitudinal: 1 mes (n=67), 3 meses (n=64), 6 meses (n=59), 9 meses (n=60), 12 meses (n=68) y 24 meses (n=70).

4.2 Medidas cognitivas / batería de pruebas

La evaluación neuropsicológica de los participantes tuvo una duración aproximada de 60-90 minutos, y se componía de una breve entrevista telefónica mediante la aplicación de la prueba TICS (Brandt et al., 1988), una entrevista presencial estructurada y una batería de 19 pruebas de evaluación neuropsicológica.

La entrevista presencial recogía datos sociodemográficos (fecha de nacimiento y edad, sexo, años de educación formal completados, estado civil, ocupación actual...) y clínicos (enfermedades relevantes, medicación...), además de otros datos de interés (consumo de tabaco y alcohol, dominancia manual, lengua materna y bilingüismo...). Las 19 pruebas de evaluación neuropsicológica incluidas en la batería de pruebas Normacog (del Pino, Peña, Schretlen, et al., 2015) se administraron en la misma sesión por una persona entrenada para ello.

4.2.1 Estudio I.

Los participantes del estudio I fueron evaluados con el test HVLT-R (Benedict et al., 1998). HVLT-R es un test breve de aprendizaje y memoria verbal que cuenta con seis versiones paralelas que permiten reducir el efecto de la práctica que tienen las pruebas de memoria. Esta prueba cuenta con 3 ensayos de aprendizaje, un ensayo a largo plazo (20-25 minutos después del 3º ensayo), y una prueba de reconocimiento.

La forma de administración y las instrucciones empleadas en este estudio fueron las incluidas en el manual original del HVLT-R (Brandt y Benedict, 2001), traducidas y adaptadas al castellano. El tiempo medio aproximado de administración fue de diez minutos (más la espera necesaria antes del 4º ensayo).

Se administró la versión 2 del HVLT-R a 444 participantes y la versión 4 a 256 participantes. Del subgrupo de 42 personas evaluadas longitudinalmente, 22 completaron en primer lugar la versión 2 del test, y 20 personas completaron primero la versión 4.

4.2.2 Estudio II.

Los participantes del estudio II fueron evaluados con el BVMT-R (Benedict et al., 1996). BVMT-R es un test breve de aprendizaje y memoria visuoespacial que cuenta con un total de seis versiones paralelas, cada una de las cuales incluye 3 ensayos de aprendizaje, un cuarto ensayo de recuerdo a largo plazo (tras 20-25 minutos de la finalización del tercero), y una prueba de reconocimiento.

La forma de administración y las instrucciones empleadas en este estudio fueron las incluidas en el manual original del BVMT-R (Benedict, 1997) traducidas y adaptadas al castellano. El tiempo medio aproximado de administración de la prueba fue de veinte minutos (más la espera necesaria antes del 4º ensayo). Se administró la versión 1 del BVMT-R a los participantes.

4.2.3 Estudio III.

Los participantes del estudio III fueron evaluados con el test UD Interferencia, un instrumento nuevo incluido en la batería de pruebas administrada para el proyecto Normacog (del Pino, Peña, Schretlen, et al., 2015) que evalúa la velocidad de procesamiento, la atención y la resistencia a la interferencia en adultos (ver Anexo 1 para consultar las instrucciones de administración).

Las características de UD Interferencia son las siguientes: a) cada lámina consta de 64 ítems (16 por cada una de las cuatro columnas); b) incluye colores fácilmente distinguibles (azul, rosa y negro); c) incluye círculos de colores en la segunda lámina; d) el tiempo de administración de cada lámina es de 30 segundos y se registra el número de ítems leídos/denominados correctamente en ese tiempo, que componen la puntuación de cada lámina; e) el índice de resistencia a la interferencia se calcula siguiendo la fórmula de Golden (Golden, 1978; Golden y Freshwater, 2002): $PC - [(Px C)/(P + C)]$; f) el orden de administración de las láminas, al igual que en algunas de las versiones más utilizadas (Comalli et al., 1962; Golden, 1978; Golden y Freshwater, 2002; Naglieri y Das, 1997), es el siguiente: 1. Palabras (P), 2. Colores (C), 3. Palabras-Colores (PC).

Además del UD Interferencia, a los participantes del estudio III, como parte de la exhaustiva evaluación neuropsicológica, se les administraron también las siguientes pruebas: TCPS (Salthouse y Babcock, 1991), Trail Making Test-A (TMT-A) y Trail Making Test-B (TMT-B) (Reitan, 1985), Brief Test of Attention (BTA) (Schretlen et al., 1996), subtest de fluidez verbal del test Calibrated Ideational Fluency Assessment (CIFA) (Schretlen y Vannorsdall, 2010) y HVLIT-R (Benedict et al., 1998).

4.2.4 Estudio IV.

Los participantes del estudio IV fueron evaluados con 16 de las 19 pruebas incluidas en la batería de pruebas Normacog (del Pino, Peña, Schretlen, et al., 2015); las 3 pruebas excluidas de estos análisis son pruebas para las cuales no se espera efecto de la

práctica por ser pruebas clínicas o funcionales (Escala de depresión geriátrica, GDS-15 y Cuestionario de actividades de la vida diaria de Lawton, AVDL) o para las cuales una variación en la puntuación no es indicativo de deterioro o mejoría (Test de estimación cognitiva).

Cada una de las 16 pruebas fue administrada a cada sujeto en dos ocasiones, en el mismo orden dentro de la batería de pruebas, y siguiendo las instrucciones de cada manual. Para las pruebas HVLT-R y BVMT-R se utilizaron versiones diferentes en la primera (HVLT-R versión 2, BVMT-R versión 1) y en la segunda evaluación (HVLT-R versión 4, BVMT-R versión 3).

Se analizaron un total de 36 variables: Test de Memoria Prospectiva (TMP); TAP; tiempo y aciertos del subtest de lectura de Pseudopalabras del PROLEC-R; MoCA; copia y memoria de la FCT; láminas P, C, PC e Índice de resistencia a la interferencia del test UD Interferencia; TCPS, con tres y seis letras; ensayo 1, recuerdo total, ensayo 4 e índice de discriminación del HVLT-R; ensayo 1, recuerdo total, ensayo 4 e índice de discriminación del BVMT-R; categorías, errores perseverativos y errores totales del M-WCST; TMT-A y TMT-B; Grooved Pegboard Test (GPT) con mano dominante y con mano no dominante; reproducción libre y copia del Test del Reloj; números y letras del BTA; Fluidez Verbal (animales, letra L, supermercado) y Test de denominación de Boston (TDB).

4.3 Análisis estadísticos

4.3.1 Análisis sociodemográficos.

Se realizaron análisis mediante la χ^2 para comparar la distribución por edad, educación y sexo entre nuestra muestra y la población española según INE y para comparar la distribución por edad y educación entre los hombres y las mujeres de nuestra muestra.

4.3.2 Normalización y estandarización.

4.3.2.1 Estudios I, II y III.

Para la normalización y estandarización de las pruebas HVLT-R, BVMT-R y UD Interferencia se realizaron los siguientes análisis estadísticos, que a su vez se encuentran detallados en el proyecto Normacog (del Pino, Peña, Schretlen, et al., 2015): a) Analizar las medias del rendimiento cognitivo por 8 rangos de edad, por 4 rangos de educación y analizar las diferencias de sexo; b) Examinar los coeficientes de correlación (r) y de determinación (r^2) de las puntuaciones directas (PD) en relación con las variables sociodemográficas edad, educación y sexo; c) Asignar los rangos de percentiles (Pc) a las PD por frecuencias acumulativas y equiparar los Pc a las puntuaciones escalares (PE) ajustadas para cada rango de edad (PE_{ae}) con un rango de 2 a 18, creando una distribución normal (media: 10 ± 3) (Jordi Peña-Casanova, Blesa, et al., 2009; Smerbeck et al., 2012; Testa et al., 2009); d) Realizar análisis de regresión múltiple para realizar la normalización ajustada por el nivel educativo, usando la siguiente fórmula: PE_n (normalizada) = $PE_{ae} (\beta \times [\text{Educación} - 12])$; se estableció como punto de corte 12 años de educación, ya que en estudios previos demostró ser un buen punto de referencia, y el resultado de la fórmula se truncó al entero inferior (Mungas et al., 1996; Jordi Peña-Casanova, Blesa, et al., 2009; Smerbeck et al., 2012).

4.3.3 Análisis de fiabilidad y validez.

4.3.3.1 Estudio III.

Se analizaron la fiabilidad y la validez concurrente y de constructo del test UD Interferencia. Para el análisis de la fiabilidad se examinaron: a) la consistencia interna de las variables P, C y PC mediante el alpha de Cronbach; b) la fiabilidad test retest mediante la correlación entre las variables en el tiempo 1 y sus homólogas en tiempo 2.

Para el análisis de la validez concurrente del UD Interferencia se analizó la correlación entre las variables del test y otros test que miden velocidad de procesamiento, atención, interferencia y/o flexibilidad: TCPS (Salthouse y Babcock, 1991), TMT-A y TMT-B (Reitan, 1985).

Para el análisis de la validez de constructo del test se calcularon cuatro puntuaciones compuestas: a) atención ($\alpha=0,82$), incluía el tiempo recodificado para llevar a cabo TMT-A y TMT-B, y la puntuación obtenida en el test BTA (Schretlen et al., 1996); b) velocidad de procesamiento ($\alpha=0,85$) incluía el número de respuestas correctas en el TCPS-3 y TCPS-6, y el tiempo recodificado para llevar a cabo la tarea en GPT (Matthews y Klove, 1964) con la mano dominante; c) fluidez verbal ($\alpha=0,75$) incluía el sumatorio de animales (1 minuto), palabras iniciadas por letra L (1 minuto) y alimentos que puedes encontrar en un supermercado (1 minuto) (Schretlen y Vannorsdall, 2010); y d) aprendizaje y memoria verbal ($\alpha=0,90$) incluía los tres ensayos de aprendizaje y el ensayo de memoria a largo plazo de HVLT-R.

4.3.4 Equivalencia entre versiones

4.3.4.1 Estudio I.

Se realizó un MANOVA de medidas repetidas de 2 (versiones) x 2 (tiempos) para analizar los factores versión, tiempo y la interacción entre los factores versión y tiempo, utilizando los datos de 42 participantes con evaluación longitudinal y con el objetivo de comprobar la equivalencia entre las versiones 2 y 4 del HVLT-R. El tiempo fue incluido como factor intrasujetos y el orden de administración de las 2 versiones como factor intersujeto.

4.3.5 Obtención de índices de cambio fiable.

4.3.5.1 Estudio IV.

Se han analizado y se proporcionan: datos descriptivos sobre las puntuaciones en tiempo 1 y tiempo 2, la diferencia entre ambas (tiempo 2 – tiempo 1), el análisis de las diferencias entre evaluaciones mediante *t* de student para muestras relacionadas, y análisis del tamaño del efecto mediante la *d* de Cohen. Se han calculado cinco índices de cambio fiable para la evaluación del cambio cognitivo, basándonos en trabajos previos (Duff, 2012, 2014; Sánchez-Benavides et al., 2016):

1) Puntuaciones de discrepancia: Se calculó la diferencia entre las puntuaciones obtenidas en el tiempo 2 y en el tiempo 1 (puntuación de discrepancia; PD) por parte de nuestra muestra. Por cada uno de los grupos longitudinales, se calculó la frecuencia acumulativa de las PD y se asignaron las puntuaciones a los percentiles más relevantes (2%, 5%, 16%, 50%, 84%, 95%, 98%).

2) Índice de Desviación Estándar (IDE): Se calculó y se proporciona la DE_{norm} de la muestra normativa en el tiempo 1 para cada una de las variables mediante la fórmula $IDE = \frac{Tiempo2-Tiempo1}{DE_{norm}}$. Se calculó el punto de corte mediante la siguiente fórmula: $DE_{norm} * \pm 1,645$.

3) Índice de Cambio Fiable (ICF): Se presenta la fórmula propuesta por Jacobson y Truax en el año 1991 (Jacobson y Truax, 1991) $ICF =$

$$\frac{Tiempo2-Tiempo1}{EED} = \frac{Tiempo2-Tiempo1}{\sqrt{2(DE_{norm}\sqrt{1-r_{xx}})^2}}$$

A partir de la DE_{norm} y el coeficiente de fiabilidad de la prueba (r_{xx}), se obtuvo el EEM, y a partir de éste se calculó el EED.

Se calculó el punto de corte mediante la siguiente fórmula: $EED * \pm 1,645$.

4) Índice de Cambio Fiable + efecto de la práctica (ICF+EP): Siguiendo la fórmula propuesta por Chelune et al. (1993) basada en el ICF, se proporciona la fórmula ICF+EP, con el EED y el efecto de la práctica (EP) (resta entre el tiempo

2 y el tiempo 1 de la muestra) calculado para cada grupo longitudinal: $ICF + EP = \frac{\text{Tiempo2} - \text{Tiempo1} - EP}{EED}$. Se calculó el punto de corte mediante la siguiente fórmula:

$$EED * \pm 1,645 + EP.$$

5) Fórmulas basadas en regresión estandarizadas: Siguiendo el trabajo de McSweeny et al. (1993), se realizaron análisis de regresión múltiple, por pasos, incluyendo la puntuación obtenida en el tiempo 1, variables sociodemográficas como la edad, los años de educación completados y el género, y los días transcurridos entre evaluaciones, como predictores de la puntuación en el tiempo 2. En la fórmula proporcionada para calcular la puntuación predicha en tiempo 2, se incluyeron únicamente las variables con influencia estadísticamente significativa. Se calculó y se proporciona el error estándar de la estimación del modelo de regresión (EEE) necesario para aplicar la fórmula:

$$\frac{\text{Tiempo2} - \text{Tiempo2 pred}}{EEE}.$$

VI. Discusión

6. Discusión

La presente tesis pretende contribuir a aumentar la disponibilidad de recursos para la evaluación neuropsicológica, proporcionando datos normativos e índices de cambio fiable adaptados a la población española que permitan una corrección adecuada de diferentes pruebas de evaluación.

El *estudio I* tenía como objetivo proporcionar normas de corrección adecuadas a las características sociodemográficas de la población española adulta para el HVLT-R, mediante la obtención de baremos estandarizados y normalizados. Con el objetivo de evaluar su influencia en el rendimiento en HVLT-R, se analizaron las variables sociodemográficas edad, educación y sexo. La edad y los años de educación tuvieron una amplia influencia y un efecto significativo en el rendimiento en HVLT-R; en línea con lo encontrado en estudios previos (Arango-Lasprilla et al., 2015; Friedman et al., 2002; Hester et al., 2004; Miotto et al., 2012; Norman et al., 2011; Ryan et al., 2020; Waldrop-Valverde et al., 2015), nuestros resultados mostraron que las personas más jóvenes y las personas con mayor educación tuvieron un rendimiento superior. Con respecto a influencia del sexo, este estudio encontró un rendimiento ligeramente superior por parte de las mujeres y sólo en algunas de las variables analizadas. Estos resultados son consistentes con los de otros estudios, ya que, pese a que las diferencias en el desempeño en aprendizaje y memoria verbal según el sexo han sido ampliamente documentadas en la bibliografía (Kimura y Clarke, 2002), la influencia del sexo en el rendimiento en el HVLT-R genera controversia; algunos estudios han encontrado que las mujeres puntúan ligeramente mejor que los hombres (Friedman et al., 2002; Norman et al., 2011; Ryan et al., 2020; Vanderploeg et al., 2000), pero otros estudios no han encontrado diferencias en cuanto al sexo (Arango-Lasprilla et al., 2015; Cherner et al., 2007; Hester et al., 2004; Miotto et al., 2012). En consecuencia, los baremos obtenidos permiten tener en cuenta la

edad y la educación de la persona evaluada, pues estas variables han demostrado tener una amplia influencia en el test. Se presentan datos normativos para cuatro variables (ensayo 1, recuerdo total, ensayo 4 e índice de discriminación); a diferencia de otros estudios normativos para la prueba HVLT-R (Arango-Lasprilla et al., 2015; Friedman et al., 2002; Hester et al., 2004; Miotto et al., 2012; Norman et al., 2011; Ryan et al., 2020; Vanderploeg et al., 2000; Waldrop-Valverde et al., 2015), este estudio presenta datos a partir de 2 versiones diferentes del test (versiones 2 y 4), que permiten readministrar la prueba evitando el aprendizaje de las palabras, una ventaja tanto para el entorno clínico como para la investigación. La evaluación del aprendizaje y la memoria verbal en población española cuenta con cierta disponibilidad pruebas (Benedet y Alejandre, 1998; Campo et al., 2000; J. Peña-Casanova, 1991), sin embargo, solamente HVLT-R cuenta con 6 versiones diferentes que permiten una evaluación breve. Este estudio aporta datos normativos y estandarizados para la prueba HVLT-R, ampliamente utilizada tanto a nivel nacional como internacional, para población española; estos datos permiten adaptar la corrección de la prueba a las características sociodemográficas de la persona evaluada, teniendo en cuenta su edad y los años de educación completados.

De forma similar al estudio I, el *estudio II* tuvo como objetivo la obtención de baremos estandarizados y normalizados para la prueba BVMT-R, proporcionando así normas de corrección adecuadas a las características sociodemográficas de la población española adulta. Para ello, se analizaron las variables sociodemográficas, con el objetivo de valorar su influencia y crear así datos normativos adecuados. Se analizó la influencia de la edad, la educación y el sexo. Nuestros resultados mostraron un efecto significativo de la edad y el nivel educativo en todas las variables analizadas, con las personas más jóvenes y/o las personas con una mayor educación, puntuando de forma superior, de forma concordante con estudios previos (Benedict et al., 1996; Duff, 2016; Kane y

Yochim, 2014; Kemmotsu et al., 2013; Miotto et al., 2012). De la misma forma que en previos estudios, estas variables tuvieron mayor influencia en los diferentes ensayos de aprendizaje y recuerdo, y menor en el índice de discriminación (Benedict et al., 1996). No se encontraron diferencias entre hombres y mujeres en ninguna de las variables analizadas; estos resultados son consistentes con los de otros estudios que indican que el sexo no tiene influencia en la memoria visuoespacial evaluada con BVMT-R (Benedict et al., 1996; Brunet et al., 2019; Corwin y Bylsma, 1993; Duff, 2016; Kane y Yochim, 2014; Kemmotsu et al., 2013; Miotto et al., 2012). Se han calculado los baremos normativos teniendo en consideración las variables con influencia en el rendimiento ante BVMT-R. Estudios de normalización previos con esta prueba ofrecen solamente datos relativos a las variables recuerdo total y ensayo 4 (Cherner et al., 2007; Gale et al., 2007; Kane y Yochim, 2014; Kemmotsu et al., 2013; Miotto et al., 2012; Norman et al., 2011), sin embargo, con el objetivo de ofrecer una información más completa, este estudio presenta datos para cuatro variables de la prueba BVMT-R (ensayo 1, recuerdo total, ensayo 4 e índice de discriminación). En población española, es frecuente la utilización de algunas pruebas para la evaluación del aprendizaje y la memoria visuoespacial (Malec et al., 1991; Osterrieth, 1944; Rey, 1941; Wechsler, 2009), sin embargo, BVMT-R destaca sobre otras pruebas por contar con 6 versiones diferentes que permiten evitar el aprendizaje de los estímulos, y por contar con 3 ensayos de aprendizaje, necesarios para permitir la aparición de una curva de aprendizaje. Este estudio aporta baremos normativos para el test BVMT-R, adaptados a las características de la población española actual; los datos normativos que se presentan permiten un ajuste individualizado según la edad y la educación de la persona evaluada.

El *estudio III* tenía dos objetivos principales, por una parte, presentar una versión del test de Stroop, el test UD Interferencia, y ofrecer sus propiedades psicométricas en

cuanto a fiabilidad y validez, y por otra, proporcionar datos normativos del test UD Interferencia, que permitan una corrección e interpretación de los datos adecuados a las características sociodemográficas de la población española adulta. UD Interferencia permite eludir algunas de las limitaciones existentes en otras versiones cuando se administran a personas con dificultades en la lectura. Por ejemplo, las personas con daltonismo y las personas mayores tienen dificultades para discriminar entre los colores verde y rojo o verde y azul, respectivamente (Pachana et al., 2004; Rivera et al., 2017; Strauss et al., 2006); por ello, UD Interferencia, mediante el uso de ítems grandes de colores distinguibles, contribuye a la disponibilidad de instrumentos de evaluación, añadiendo una nueva herramienta a otras que han remediado parcialmente este problema (Pachana et al., 2004). Además, el reducido tiempo de administración de este test facilita su inclusión en una batería de pruebas y su utilización con pacientes fácilmente fatigables. El test UD Interferencia ha demostrado tener una alta fiabilidad test-retest para cada una de las 3 láminas administradas y muy buena consistencia interna. Además, este test cuenta con una alta validez concurrente, ya que existe una alta relación entre UD Interferencia y otros test de referencia en la medición de la velocidad de procesamiento y la atención, en línea con estudios anteriores, que han demostrado que otras versiones del test de Stroop también se relacionan con test que evalúan estos dominios cognitivos (Strauss et al., 2006). Los constructos velocidad psicomotora y atención tienen una mejor correlación con UD Interferencia que los constructos fluidez verbal y aprendizaje y memoria verbal, lo que sugiere una buena validez de constructo. Previo a la creación de los baremos normativos se analizaron las variables sociodemográficas. La edad y el nivel educativo influyeron de forma significativa en el rendimiento de UD Interferencia, de manera que a mayor edad y/o menor nivel educativo, se obtuvo un peor rendimiento, en concordancia con estudios previos con otras versiones del Test de Stroop (Moering et al., 2004; Pachana

et al., 2004; Jordi Peña-Casanova, Quiñones-Úbeda, et al., 2009; Rivera et al., 2015; Rognoni et al., 2013; Steinberg et al., 2005; Strauss et al., 2006; Troyer et al., 2006; van Boxtel et al., 2001; Van der Elst et al., 2006). El sexo, sin embargo, sólo mostró diferencias significativas en el índice de resistencia a la interferencia, con una puntuación ligeramente superior por parte de los hombres; la ausencia de diferencias entre sexos en la puntuación de las láminas del test ya había sido reflejada en la literatura (Jordi Peña-Casanova, Quiñones-Úbeda, et al., 2009; Rivera et al., 2015; Troyer et al., 2006). La forma óptima de comparar el rendimiento en las diferentes láminas es utilizar datos normativos como la PE o el Pc, ya que la dificultad entre láminas varía (Golden, 2020). Teniendo en cuenta la influencia de las variables sociodemográficas, se han creado datos normativos para cada una de las tres láminas incluidas en el test y para el índice de resistencia a la interferencia, sin embargo, este último debe interpretarse con precaución (Chafetz y Matthews, 2004), por ser una medida variable en el tiempo. Este estudio aporta el test UD Interferencia, una herramienta útil por sus características que permite su inclusión en una batería más larga de evaluación por su reducido tiempo de administración, y datos normativos para población española que permiten adaptar su corrección teniendo en cuenta la edad y el nivel educativo de la persona evaluada.

El *estudio IV* tenía como objetivo proporcionar índices de cambio fiable para 16 pruebas de evaluación neuropsicológica, mediante cinco métodos (percentiles de las puntuaciones de discrepancia, IDE, ICF, ICF+EP, y fórmulas basadas en regresión estandarizadas). Estos datos, adecuados a seis tiempos longitudinales diferentes (1 mes, 3 meses, 6 meses, 9 meses, 12 meses y 24 meses), nos permiten estimar un rango de cambio normal y esperable que una persona podría tener entre dos evaluaciones, y por tanto, si un cambio cognitivo se puede considerar o no fiable. La existencia del efecto de la práctica se hace patente, ya que, en la segunda evaluación en los grupos de 1 a 12 meses

las puntuaciones son significativamente mejores para la mayoría de variables, en cambio, en el grupo de 24 meses, son mayoría las variables que no muestran diferencias significativas entre evaluaciones; estudios previos han encontrado que la duración del intervalo entre evaluaciones influye en la magnitud del efecto de la práctica (McCaffrey et al., 2000), que puede encontrarse en lapsos de tiempo desde una semana hasta dos años y medio (Calamia et al., 2012; Duff, 2014; Duff et al., 2005; Levine et al., 2004; Salthouse, 2010; Sánchez-Benavides et al., 2016). Este efecto influye sobre todo en las pruebas de memoria (Nieto et al., 2014; Sánchez-Benavides et al., 2016); en la misma línea, nuestros resultados muestran que las variables más afectadas por el efecto de la práctica corresponden a test de memoria, pruebas susceptibles de ser memorizadas por recordarse la forma correcta de ejecución como el M-WCST, o por estar compuestas por palabras fácilmente memorizables como el TDB – 15. WCST se ha definido como una “prueba de un solo disparo”, considerando su readministración poco fiable cuando la persona ya lo ha resuelto previamente (Lezak et al., 2012; Strauss et al., 2006), por lo que, pese a la obtención de datos sobre el cambio fiable, recomendamos interpretar con prudencia una mejoría entre evaluaciones. En línea con nuestros resultados, otros estudios han encontrado diferencias atribuibles al efecto de la práctica cuando el TDB se ha readministrado entre 9 y 15 meses después de la primera evaluación (Sachs et al., 2012; Zec et al., 2005), por lo que se aconseja utilizar las cuatro versiones reducidas de este test, que nos permiten una administración rápida evitando el efecto de la práctica (Kent y Luszcz, 2002; Zgaljardic y Benedict, 2001). Asimismo, se recomienda el uso de pruebas paralelas también para las pruebas HVLT-R y BVMT-R, de la misma forma que se ha hecho en este estudio, que únicamente ha encontrado una ligera mejoría en BVMT-R, en concordancia con estudios previos de sus autores (Benedict y Zgaljardic, 1998). Aunque otros estudios evitaron el uso de versiones alternativas en estas pruebas (Duff, 2014), en

este estudio se han utilizado diferentes versiones con el objetivo de proporcionar datos adecuados para la práctica clínica. Existen diferentes predictores de la puntuación en el segundo tiempo; la puntuación en el primer tiempo ha sido el predictor más influyente en el modelo de regresión, en línea con estudios previos, que argumentan que “el comportamiento pasado es el mejor predictor del comportamiento futuro” (Attix et al., 2009; Sánchez-Benavides et al., 2016). La edad y los años de educación han demostrado ser características relevantes e influyentes con respecto al cambio esperado en la segunda evaluación; se espera una mejoría mayor cuanto menor es la edad y/o mayor la educación del evaluado, en concordancia con estudios previos (Attix et al., 2009; Calamia et al., 2012; Duff, 2014; Duff et al., 2005; Sánchez-Benavides et al., 2016). A este respecto, es importante contar con índices de cambio fiable que nos permitan valorar si el cambio se ajusta a lo esperable o no, ya que algunos test pueden estar más influidos por los cambios asociados a la edad que por el efecto de la práctica (Sánchez-Benavides et al., 2016). El sexo sólo ha mostrado su influencia en tres de las variables analizadas; investigaciones previas apoyan la escasa influencia del sexo en la predicción de la puntuación en la segunda evaluación (Duff, 2014; Duff et al., 2004, 2005; Sánchez-Benavides et al., 2016). Tal y como ocurre en otros estudios (Duff et al., 2004, 2005; Levine et al., 2004), el intervalo entre evaluaciones ha demostrado influir significativamente sólo en algunas variables. Los cinco índices propuestos para la evaluación del cambio fiable son métodos ampliamente utilizados en la literatura (Duff, 2012; Hinton-Bayre, 2010). No existe consenso con respecto a cuál de los métodos es el mejor, aunque se ha observado que el ICF+EP y los métodos basados en regresión son los que mayor concordancia tienen entre ellos y los que mejor identifican el cambio (Frerichs y Tuokko, 2005; Heaton et al., 2001; Temkin et al., 1999), posiblemente porque estos métodos permiten tener en cuenta el efecto de la práctica (Maassen et al., 2009). Es necesaria la obtención de índices de

cambio fiable para población española, hasta ahora escasos; este estudio contribuye con una herramienta muy necesaria para la práctica de la neuropsicología clínica, aportando datos para población mayor de 18 años y para pruebas ampliamente utilizadas tanto a nivel nacional como internacional, sumándose a un estudio previo que proporciona datos para personas mayores de 50 años (Sánchez-Benavides et al., 2016). Este estudio aporta por primera vez cinco índices de cambio fiable adaptados a seis momentos temporales diferentes, para población española, que permiten valorar cuando un cambio entre evaluaciones es significativo.

Los cuatros estudios presentados anteriormente comparten algunas limitaciones, relacionadas con la obtención de la muestra que compone los estudios. Como es común en los estudios normativos, los datos están limitados a las características de las personas evaluadas; por ello, se ha tratado de seleccionar una muestra suficientemente representativa de la población española, teniendo en cuenta los datos publicados por el INE. Pese a ello, las personas con niveles educativos superiores tienen una mayor representatividad; esto ocurre sobre todo en los grupos de edad más jóvenes, y en menor medida en los grupos de edad más avanzados para los cuales es difícil acceder a personas con alto nivel educativo, algo representativo de las personas mayores en España (Montoya-Murillo et al., 2019), ya que éstas tuvieron difícil acceso a la educación por la situación que atravesaba el país durante su edad escolar (Arco Blanco, 2006). Aun así, los estudios de normalización incluidos en la presente tesis, proporcionan datos para corregir la prueba teniendo en cuenta la edad y la educación del examinado, por lo que la corrección permite obtener una puntuación normalizada ajustada a cada persona de manera individualizada. Las normas de corrección incluidas en el presente estudio se han construido mediante la técnica basada en frecuencia acumulativa, ofreciendo así unos baremos más sencillos de utilizar en la práctica clínica que las normas basadas en

regresión, y más ampliamente utilizados en los test de evaluación neuropsicológica. Teniendo en cuenta que la muestra que compone el estudio es una muestra que ha realizado su educación en España, no podemos garantizar que los datos sean representativos si se utilizan en la evaluación de personas cuya educación se haya realizado en otros países; aunque como estudios previos mencionan, estos datos podrían ser de interés para otras poblaciones, en ausencia de datos más adecuados (Sánchez-Benavides et al., 2016).

En conclusión, estos cuatro estudios pretenden aumentar la disponibilidad de herramientas para la adecuada evaluación neuropsicológica para población española, proporcionando: baremos normativos y estandarizados para los test de aprendizaje y memoria HVLT-R y BVMT-R y para el test UD Interferencia, una nueva versión del test de Stroop, que permiten el ajuste individualizado según la edad y educación de cada persona; y cinco índices de cambio fiable para estimar la magnitud del cambio entre evaluaciones sucesivas, que permitirán una mejor interpretación de los resultados, evitando algunos sesgos y errores de interpretación comunes en las evaluaciones sucesivas.

VII. Conclusiones

7. Conclusiones

Las principales conclusiones de esta tesis, derivadas de los cuatro estudios descritos, pueden resumirse de la siguiente manera:

- Los datos normativos proporcionados para los test HVLIT-R, BVMT-R y UD Interferencia permiten su ajuste tanto por la edad como por el nivel educativo de cada persona, lo que aporta una puntuación normalizada ajustada a cada persona de manera individualizada y quizás por ello, una herramienta más fiable.
- El test UD Interferencia mostró buenas características psicométricas: concretamente mostró una alta fiabilidad test-retest para cada una de las tres láminas administradas y muy buena consistencia interna, además de una alta validez concurrente y de constructo. Este test permite la evaluación de la velocidad de procesamiento, la atención, y la resistencia a la interferencia en una población más amplia que otras versiones del Test de Stroop, ya que incluye colores fácilmente distinguibles, un tamaño de letra más grande y un reducido tiempo de administración.
- Los cinco índices de cambio fiable aportados permiten evaluar la magnitud del cambio observado entre dos evaluaciones sucesivas para 16 pruebas de evaluación neuropsicológica, contribuyendo así a una mejor interpretación de los resultados obtenidos en situaciones donde el efecto de la práctica o de regresión a la media tienen un rol significativo. Estos datos servirán de apoyo tanto en futuras investigaciones como en la evaluación clínica y el diagnóstico de diversas patologías, ya que permitirán tanto a clínicos como a investigadores evitar algunos de los sesgos y errores de interpretación habituales cuando se evalúa el cambio entre evaluaciones sucesivas.

Conclusions

The main conclusions of this thesis, derived from the four studies described, can be summarized as follows:

- Normative data provided for HVLT-R, BVMT-R and UD Interference tests allow the adjustment by the age and educational level of the person. It provides a standardized score adjusted to each person individually and perhaps, therefore, a more reliable tool.
- The UD Interference test showed good psychometric characteristics: specifically, it showed a high test-retest reliability for each of the three administered pages and a very good internal consistency, as well as high concurrent and construct validity. This test allows the assessment of processing speed, attention and resistance to interference in a wider population than other versions of the Stroop Test, as it includes easily distinguishable colors, a larger font size and a reduced administration time.
- The five reliable change indices provided, allow to assess the magnitude of the change observed between two successive evaluations for 16 neuropsychological assessment tests, helping to a better interpretation of the obtained results in situations where the practice effect or regression to the mean play a significant role. These data will support both future research and clinical assessment and in the diagnosis of various illnesses, since they will allow both clinicians and researchers to avoid some of common biases and errors of interpretation when assessing change in repeated assessments.

Ondorioak

Tesi honen ondorio nagusiak, deskribatutako lau ikerketetatik eratorritakoak, honela labur daitezke:

- HVLТ-R, BVMT-R eta UD Interferentzia probetarako emandako datu normatiboek pertsonaren adina eta hezkuntza mailari egokitzea ahalbidetzen dute, pertsona bakoitzari egokitutako puntuazio estandarizatuak eskainiz, eta agian, horregatik, tresna fidagarriagoa eskainiz.
- UD Interferentzia probak ezaugarri psikometriko onak erakutsi zituen: zehazki, administratutako hiru orrialdeetarako test-retest fidagarritasun handia eta barneko konsistentzia oso ona erakutsi zuen, bai eta baterako-baliozkotasun eta konstruktibaliokotasun handia ere. Proba honek prozesamenduaren abiadura, arreta eta interferentziarekiko erresistentzia ebaluatzea ahalbidetzen du Stroop Testaren beste bertsioek baino populazio zabalagoan, aintzat hartuz erraz bereizten diren koloreak, letra tamaina handiagoa eta administrazio denbora murriztua.
- Emandako bost aldaketa-indize fidagarriek segidako ebaluazioen artean ikusitako aldaketaren magnitudea ebaluatzea ahalbidetzen dute, 16 ebaluazio froga neuropsikologikorako, lortutako emaitzen interpretazio hobea lortuz praktikaren efektua edo batez bestekoarekiko itzulera garrantzi nabarmena duten egoeretan. Datu horiek etorkizuneko ikerketen, ebaluazio klinikoan eta hainbat gaixotasunen diagnostikoan lagunduko dute, izan ere, kliniko eta ikertzaileei segidako ebaluazioen arteko aldaketa ebaluatzerakoan isuri eta interpretazio akats ohikoenak saihesteko aukera emango bait die.

VIII. Referencias

8. Referencias

- Albert, M., Moss, M. B., Blacker, D., Tanzi, R. y McArdle, J. J. (2007). Longitudinal change in cognitive performance among individuals with mild cognitive impairment. *Neuropsychology*, *21*(2), 158-169. <https://doi.org/10.1037/0894-4105.21.2.158>
- Anastasi, A. y Urbina, S. (1997). *Psychological testing (7th ed.)*. Prentice Hall/Pearson Education.
- Arango-Lasprilla, J. C., Rivera, D., Garza, M. T., Saracho, C. P., Rodríguez, W., Rodríguez-Agudelo, Y., Aguayo, A., Schebela, S., Luna, M., Longoni, M., Martínez, C., Doyle, S., Ocampo-Barba, N., Galarza-Del-Angel, J., Aliaga, A., Bringas, M., Esenarro, L., García-Egan, P. y Perrin, P. B. (2015). Hopkins Verbal Learning Test- Revised: Normative data for the Latin American Spanish speaking adult population. *NeuroRehabilitation*, *37*(4), 699-718. <https://doi.org/10.3233/NRE-151286>
- Arco Blanco, M. Á. del. (2006). «Morir de hambre»: autarquía, escasez y enfermedad en la España del primer franquismo. *Pasado y memoria*, *5*, 241-258. <https://doi.org/10.14198/pasado2006.5.12>
- Ardila, A., Ostrosky-Solis, F., Rosselli, M. y Gomez, C. (2000). Age-Related Cognitive Decline During Normal Aging: The Complex Effect of Education. *Archives of Clinical Neuropsychology*, *15*(6), 495-513. <https://doi.org/10.1093/arclin/15.6.495>
- Ardila, A. y Rosselli, M. (2007). *Neuropsicología clínica*. Manual Moderno.
- Argento, O., Smerbeck, A., Pisani, V., Magistrale, G., Incerti, C. C., Caltagirone, C., Benedict, R. H. B., Nocentini, U., Argento, O., Smerbeck, A., Pisani, V.,

- Magistrale, G., Incerti, C. C., Caltagirone, C., Benedict, R. H. B. y Nocentini, U. (2016). Regression-Based Norms for the Brief Visuospatial Memory Test-Revised in Italian population and application in MS patients. *The Clinical Neuropsychologist*, 30(51), 1469-1478.
<https://doi.org/10.1080/13854046.2016.1183713>
- Arnett, P. y Forn Frias, C. (2007). Evaluación neuropsicológica en la esclerosis múltiple. *Revista de Neurología*, 44(03), 166.
<https://doi.org/10.33588/rn.4403.2006057>
- Attix, D. K., Story, T. J., Chelune, G. J., Ball, J. D., Stutts, M. L., Hart, R. P. y Barth, J. T. (2009). The prediction of Change: Normative neuropsychological trajectories. *The Clinical Neuropsychologist*, 23(1), 21-38.
<https://doi.org/10.1080/13854040801945078>
- Barnett, A. G. (2004). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology*, 34(1), 215-220.
<https://doi.org/10.1093/ije/dyh299>
- Benedet, M. y Alejandre, M. (1998). *TAVEC: test de aprendizaje verbal España-Complutense*. TEA. <https://doi.org/http://dx.doi.org/10.5093/cl2013a18>
- Benedict, R. H. B. (1988). Group Format Memory Retraining: A program designed for an empirical evaluation of the effectiveness of memory retraining. *Paper presented at the seventh annual convention of the National Head-Injury Foundation, Atlanta, GA.*
- Benedict, R. H. B. (1997). *Brief Visuospatial Memory Test-Revised. Professional manual*. Psychological Assessment Resources Inc.
- Benedict, R. H. B. y Brandt, J. (2007). *HVLT-R/BVMT-R professional manual*

supplement. Psychological Assessment Resources Inc.

Benedict, R. H. B., Cookfair, D., Gavett, R., Gunther, M., Munschauer, F., Garg, N. y

Weinstock-Guttman, B. (2006). Validity of the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). *Journal of the International Neuropsychological Society : JINS*, 12(4), 549-558.

<https://doi.org/10.1017/S1355617706060723>

Benedict, R. H. B., Schretlen, D., Groninger, L. y Brandt, J. (1998). Hopkins Verbal Learning Test – Revised: Normative Data and Analysis of Inter-Form and Test-Retest Reliability. *The Clinical Neuropsychologist*, 12(1), 43-55.

<https://doi.org/10.1076/clin.12.1.43.1726>

Benedict, R. H. B., Schretlen, D. J., Groninger, L., Dobraski, M. y Shpritz, B. (1996).

Revision of the Brief Visuospatial Memory Test: Studies of Normal Performance, Reliability, and Validity. *Psychological Assessment*, 8(2), 145-153.

<https://doi.org/10.1037/1040-3590.8.2.145>

Benedict, R. H. B. y Zgaljardic, D. J. (1998). Practice Effects During Repeated

Administrations of Memory Tests With and Without Alternate Forms. *Journal of Clinical and Experimental Neuropsychology*, 20(3), 339-352.

<https://doi.org/10.1076/jcen.20.3.339.822>

Bird, C. M., Papadopoulou, K., Ricciardelli, P., Rossor, M. N. y Cipolotti, L. (2003).

Test-retest reliability, practice effects and reliable change indices for the recognition memory test. *British Journal of Clinical Psychology*, 42(4), 407-425.

<https://doi.org/10.1348/014466503322528946>

Brandt, J. y Benedict, R. H. B. (2001). *Hopkins Verbal Learning Test–Revised*.

Professional manual. Psychological Assessment Resources Inc.

- Brandt, J., Spencer, M. y Folstein, M. (1988). The telephone interview for cognitive status. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 1(2), 111-117.
- Brunet, H. E., Caldwell, J. Z. K., Brandt, J. y Miller, J. B. (2019). Influence of sex differences in interpreting learning and memory within a clinical sample of older adults. *Aging, Neuropsychology, and Cognition*, 27(1), 18-39.
<https://doi.org/10.1080/13825585.2019.1566433>
- Calamia, M., Markon, K. y Tranel, D. (2012). Scoring Higher the Second Time Around: Meta-Analyses of Practice Effects in Neuropsychological Assessment. *The Clinical Neuropsychologist*, 26(4), 543-570.
<https://doi.org/10.1080/13854046.2012.680913>
- Campo, P., Morales, M. y Juan-Malpartida, M. (2000). Versiones españolas del test de Recuerdo Verbal Selectivo. *Psicothema*, 12(2), 108-110.
- Chafetz, M. D. y Matthews, L. H. (2004). A New interference score for the Stroop test. *Archives of Clinical Neuropsychology*, 19(4), 555-567.
<https://doi.org/10.1016/j.acn.2003.08.004>
- Chelune, G. J., Naugle, R. I., Lüders, H., Sedlak, J. y et al. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology*, 7(1), 41-52. <https://doi.org/10.1037/0894-4105.7.1.41>
- Cherner, M., Suarez, P., Lazzaretto, D., Artiola i Fortuny, L., Rivera Mindt, M., Dawes, S., Marcotte, T., Grant, I. y Heaton, R. (2007). Demographically corrected norms for the Brief Visuospatial Memory Test-revised and Hopkins Verbal Learning Test-revised in monolingual Spanish speakers from the U.S.–Mexico border region. *Archives of Clinical Neuropsychology*, 22(3), 343-353.

<https://doi.org/10.1016/j.acn.2007.01.009>

Comalli, P. E., Wapner, S. y Werner, H. (1962). Interference Effects of Stroop Color-Word Test in Childhood, Adulthood, and Aging. *The Journal of Genetic Psychology*, 100(1), 47-53. <https://doi.org/10.1080/00221325.1962.10533572>

Corwin, J. y Bylsma, F. W. (1993). Psychological examination of traumatic

encephalopathy. *Clinical Neuropsychologist*, 7(1), 3-21.

<https://doi.org/10.1080/13854049308401883>

Cuetos, F., Rodríguez, B., Ruano, E. y Arribas, D. (2014). *PROLEC-R. Batería de evaluación de los procesos lectores, revisada (5ª edición)* (T. Ediciones (Ed.)).

Dean, J. A., Coulombier, D., Burton, A. H., Brendel, K. A., Smith, D. C., Dicker, R. C., Sullivan, K. M., Fagan, R. F., Dean, J. A., Coulombier, D., Smith, D. C., Brendel, K. A., Arner, T. G., Dean, A. G. y Dean, A. G. (1996). Epi Info, version 6: a word-processing, database, and statistics program for public health on IBM-compatible microcomputers. En *Atlanta: Centers for Disease Control and Prevention*.

del Pino, R., Peña, J., Ibarretxe-Bilbao, N., Schretlen, D. J. y Ojeda, N. (2015). Test de la figura compleja de Taylor: administración y corrección según un proceso de normalización y estandarización en población española. *Revista de Neurología*, 61(09), 395. <https://doi.org/10.33588/rn.6109.2015168>

del Pino, R., Peña, J., Ibarretxe-Bilbao, N., Schretlen, D. J. y Ojeda, N. (2016). Test modificado de clasificación de tarjetas de Wisconsin: normalización y estandarización de la prueba en población española. *Revista de Neurología*, 62(05), 193. <https://doi.org/10.33588/rn.6205.2015274>

del Pino, R., Peña, J., Ibarretxe-Bilbao, N., Schretlen, D. J. y Ojeda, N. (2018).

Demographically Calibrated Norms for Two Premorbid Intelligence Measures:

- The Word Accentuation Test and Pseudo-Words Reading Subtest. *Frontiers in Psychology*, 9, 1950. <https://doi.org/10.3389/fpsyg.2018.01950>
- del Pino, R., Peña, J., Schretlen, D. J., Ibarretxe-Bilbao, N. y Ojeda, N. (2015). Estudio multicéntrico de normalización y estandarización de instrumentos neuropsicológicos en personas sanas para población española: metodología y características del proyecto Normacog. *Revista de Neurología*, 61(02), 57. <https://doi.org/10.33588/rn.6102.2014481>
- Del Ser, T., González-Montalvo, J.-I., Martínez-Espinosa, S., Delgado-Villapalos, C. y Bermejo, F. (1997). Estimation of Premorbid Intelligence in Spanish People with the Word Accentuation Test and Its Application to the Diagnosis of Dementia. *Brain and cognition*, 33, 343-356. <https://doi.org/10.1006/brcg.1997.0877>
- Delaney, R. C., Prevey, M. L., Cramer, J., Mattson, R. H. y V. E. C. S. R. Group. (1992). Test-retest comparability and control subject data for the rey-auditory verbal learning test and rey-osterrieth/taylor complex figures,. *Archives of Clinical Neuropsychology*, 7(6), 523-528. [https://doi.org/10.1016/0887-6177\(92\)90142-A](https://doi.org/10.1016/0887-6177(92)90142-A)
- Duff, K. (2012). Evidence-Based Indicators of Neuropsychological Change in the Individual Patient: Relevant Concepts and Methods. *Archives of Clinical Neuropsychology*, 27(3), 248-261. <https://doi.org/10.1093/arclin/acr120>
- Duff, K. (2014). One-week practice effects in older adults: Tools for assessing cognitive change. *Clinical Neuropsychologist*. <https://doi.org/10.1080/13854046.2014.920923>
- Duff, K. (2016). Demographically corrected normative data for the Hopkins Verbal Learning Test-Revised and Brief Visuospatial Memory Test-Revised in an elderly sample. *Applied Neuropsychology: Adult*, 23(3), 179-185.

<https://doi.org/10.1080/23279095.2016.1030019>

Duff, K., Lyketsos, C. G., Beglinger, L. J., Chelune, G., Moser, D. J., Arndt, S., Schultz, S. K., Paulsen, J. S., Petersen, R. C. y Mccaffrey, R. J. (2011). Practice effects predict cognitive outcome in amnesic Mild Cognitive Impairment. *Am J Geriatr Psychiatry.*, 19(11), 932-939.

<https://doi.org/10.1097/JGP.0b013e318209dd3a.Practice>

Duff, K., Schoenberg, M. R., Patton, D., Mold, J., Scott, J. G. y Adams, R. L. (2004). Predicting change with the RBANS in a community dwelling elderly sample. *Journal of the International Neuropsychological Society*, 10(6), 828-834.

<https://doi.org/10.1017/S1355617704106048>

Duff, K., Schoenberg, M. R., Patton, D., Paulsen, J. S., Bayless, J. D., Mold, J., Scott, J. G. y Adams, R. L. (2005). Regression-based formulas for predicting change in RBANS subtests with older adults. *Archives of Clinical Neuropsychology*.

<https://doi.org/10.1016/j.acn.2004.07.007>

Frank, R. M. y Byrne, G. J. (2000). The clinical utility of the Hopkins Verbal Learning Test as a screening test for mild dementia. *International Journal of Geriatric Psychiatry*, 15(4), 317-324. [https://doi.org/10.1002/\(SICI\)1099-1166\(200004\)15:4<317::AID-GPS116>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1099-1166(200004)15:4<317::AID-GPS116>3.0.CO;2-7)

Frerichs, R. y Tuokko, H. (2005). A comparison of methods for measuring cognitive change in older adults. *Archives of Clinical Neuropsychology*, 20(3), 321-333.

<https://doi.org/10.1016/j.acn.2004.08.002>

Friedman, M. A., Schinka, J. A., Mortimer, J. A. y Graves, A. B. (2002). Hopkins verbal learning test-revised: Norms for elderly African Americans. *The Clinical Neuropsychologist*, 16(3), 356-372. <https://doi.org/10.1076/clin.16.3.356.13857>

Gale, S. D., Baxter, L., Connor, D. J., Herring, A. y Comer, J. (2007). Sex differences on the Rey Auditory Verbal Learning Test and the Brief Visuospatial Memory Test – Revised in the elderly : Normative data in 172 participants. *Journal of Clinical and Experimental Neuropsychology*, 29(5), 561–567.
<https://doi.org/10.1080/13803390600864760>

Gale, S. D., Baxter, L. y Thompson, J. (2016). Greater Memory Impairment in Dementing Females than Males Relative to Sex Matched Healthy Controls. *J Clin Exp Neuropsychol*, 38(5), 527-533.
<https://doi.org/10.1080/13803395.2015.1132298>.Greater

Gil, R. (2019). *Neuropsicología (7 ed.)*. Elsevier.

Golden, C. J. (1978). *Stroop Color and Word Test: A manual for clinical and experimental uses*. Stoelting.

Golden, C. J. (2020). *STROOP. Test de Colores y Palabras – Edición Revisada (B. Ruiz-Fernández, T. Luque y F. Sánchez-Sánchez, adaptadores)*. TEA Ediciones.

Golden, C. J. y Freshwater, S. M. (2002). *Stroop Color and Word Test: Revised examiner's manual*. Stoelting.

Gómez-Gastiasoro, A., Peña, J., Ibarretxe-Bilbao, N., Lucas-Jiménez, O., Díez-Cirarda, M., Rilo, O., Montoya-Murillo, G., Zubiaurre-Elorza, L. y Ojeda, N. (2019). A Neuropsychological Rehabilitation Program for Cognitive Impairment in Psychiatric and Neurological Conditions: A Review That Supports Its Efficacy. *Behavioural Neurology*, 2019, 1-11. <https://doi.org/10.1155/2019/4647134>

Gómez-Gastiasoro, A., Zubiaurre-Elorza, L., Peña, J., Ibarretxe-Bilbao, N., Rilo, O., Schretlen, D. J. y Ojeda, N. (2019). Altered frontal white matter asymmetry and its implications for cognition in schizophrenia : A tractography study. *NeuroImage*:

Clinical, 22(101781). <https://doi.org/10.1016/j.nicl.2019.101781>

- Gonzalez-Palau, F., Franco, M., Jimenez, F., Parra, E., Bernate, M. y Solis, A. (2013). Clinical Utility of the Hopkins Verbal Test-Revised for Detecting Alzheimer's Disease and Mild Cognitive Impairment in Spanish Population. *Archives of Clinical Neuropsychology*, 28(3), 245-253. <https://doi.org/10.1093/arclin/act004>
- Grau-Guinea, L., Pérez-Enríquez, C., García-Escobar, G., Arrondo-Elizarán, C., Pereira-Cutiño, B., Florido-Santiago, M., Piqué-Candini, J., Planas, A., Paez, M., Peña-Casanova, J. y Sánchez-Benavides, G. (2018). Desarrollo, estudio de equivalencia y datos normativos de la versión española B del Free and Cued Selective Reminding Test. *Neurología*. <https://doi.org/10.1016/j.nrl.2018.02.002>
- Hawkins, K. A., Dean, D. y Pearlson, G. D. (2004). Alternative forms of the Rey Auditory Verbal Learning Test: a review. *Behavioural neurology*, 15(3-4), 99-107. <https://doi.org/10.1155/2004/940191>
- Heaton, R. K., Taylor, M. J. y Manly, J. (2003). Demographic Effects and Use of Demographically Corrected Norms with the WAIS-III and WMS-III. En *Clinical Interpretation of the WAIS-III and WMS-III* (pp. 181-210). Elsevier. <https://doi.org/10.1016/B978-012703570-3/50010-9>
- Heaton, R. K., Temkin, N. R., Dikmen, S., Avitable, N., Taylor, M. J., Marcotte, T. D. y Grant, I. (2001). Detecting change: A comparison of three neuropsychological methods, using normal and clinical samples. *Archives of Clinical Neuropsychology*, 16(1), 75-91. <https://doi.org/10.1093/arclin/16.1.75>
- Heilbronner, R. L., Sweet, J. J., Attix, D. K., Krull, K. R., Henry, G. K. y Hart, R. P. (2010). Official position of the American Academy of Clinical Neuropsychology on serial neuropsychological assessments: the utility and challenges of repeat test

- administrations in clinical and forensic contexts. *The Clinical Neuropsychologist*, 24(8), 1267-1278. <https://doi.org/10.1080/13854046.2010.526785>
- Hester, R. L., Kinsella, G. J., Ong, B. y Turner, M. (2004). Hopkins Verbal Learning Test: Normative data for older Australian adults. *Australian Psychologist*, 39(3), 251-255. <https://doi.org/10.1080/00050060412331295063>
- Hinton-Bayre, A. D. (2010). Deriving Reliable Change Statistics from Test-Retest Normative Data: Comparison of Models and Mathematical Expressions. *Archives of Clinical Neuropsychology*, 25(3), 244-256. <https://doi.org/10.1093/arclin/acq008>
- Hogervorst, E., Combrinck, M., Lapuerta, P., Rue, J., Swales, K. y Budge, M. (2002). The Hopkins Verbal Learning Test and Screening for Dementia. *Dementia and Geriatric Cognitive Disorders*, 13(1), 13-20. <https://doi.org/10.1159/000048628>
- Iraurgi, I. (2009). Evaluación de resultados clínicos I: Entre la significación estadística y la relevancia clínica. *Norte de Salud Mental*, 8(33), 94-108.
- Jacobson, N. S. y Truax, P. (1991). Clinical significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research. *Journal of Consulting and Clinical Psychology*, 59(1), 12-19. <https://doi.org/10.1037/0022-006X.59.1.12>
- Kane, K. D. y Yochim, B. P. (2014). Construct Validity and Extended Normative Data for Older Adults for the Brief Visuospatial Memory Test , Revised. *American Journal of Alzheimer's Disease & Other Dementias*, 29(7), 601-606. <https://doi.org/10.1177/1533317514524812>
- Kazdin, A. E. (1998). *Research design in clinical psychology (3rd ed.)*. Allyn & Bacon.
- Kemmotsu, N., Enobi, Y. y Murphy, C. (2013). Performance of older Japanese American adults on selected cognitive instruments. *Journal of the International Neuropsychological Society*, 19(7), 773-781.

<https://doi.org/10.1017/S1355617713000520>

Kent, P. S. y Luszcz, M. A. (2002). A Review of the Boston Naming Test and Multiple-Occasion Normative Data for Older Adults on 15-Item Versions. *The Clinical Neuropsychologist*, *16*(4), 555-574. <https://doi.org/10.1076/clin.16.4.555.13916>

Kimura, D. (1999). *Sex and cognition*. MIT press.

Kimura, D. y Clarke, P. G. (2002). Women's Advantage on Verbal Memory is Not Restricted to Concrete Words. *Psychological Reports*, *91*(3), 1137-1142.

Kolb, B. y Wishaw, I. Q. (2015). *Fundamentals of Human Neuropsychology (7th ed.)*. Worth Publishers.

Langdon, D. W., Amato, M. P., Boringa, J., Brochet, B., Foley, F., Fredrikson, S., Hämäläinen, P., Hartung, H. P., Krupp, L., Penner, I. K., Reder, A. T. y Benedict, R. H. B. (2012). Recommendations for a brief international cognitive assessment for multiple sclerosis (BICAMS). *Multiple Sclerosis Journal*, *18*(6), 891-898. <https://doi.org/10.1177/1352458511431076>

Levine, A. J., Miller, E. N., Becker, J. T., Selnes, O. A. y Cohen, B. A. (2004). Normative Data for Determining Significance of Test–Retest Differences on Eight Common Neuropsychological Instruments. *The Clinical Neuropsychologist*, *18*(3), 373-384. <https://doi.org/10.1080/1385404049052420>

Lezak, M. D., Howieson, D. B., Tranel, E. D. y Bigler, D. (2012). *Neuropsychological assessment, 5th edition*. Oxford University Press.

Lim, Y. Y., Prang, K. H., Cysique, L., Pietrzak, R. H., Snyder, P. J. y Maruff, P. (2009). A method for cross-cultural adaptation of a verbal memory assessment. *Behavior Research Methods*, *41*(4), 1190-1200. <https://doi.org/10.3758/BRM.41.4.1190>

- Lucas-Jiménez, O., Díez-Cirarda, M., Ojeda, N., Peña, J., Cabrera-Zubizarreta, A. y Ibarretxe-Bilbao, N. (2015). Verbal Memory in Parkinson's Disease: A Combined DTI and fMRI Study. *Journal of Parkinson's Disease*, 5(4), 793-804.
<https://doi.org/10.3233/JPD-150623>
- Maassen, G. H., Bossema, E. y Brand, N. (2009). Reliable change and practice effects: Outcomes of various indices compared. *Journal of Clinical and Experimental Neuropsychology*, 31(3), 339-352. <https://doi.org/10.1080/13803390802169059>
- Malec, J. F., Ivnik, R. J. y Hinkeldey, N. S. (1991). Visual Spatial Learning Test. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3(1), 82-88. <https://doi.org/10.1037/1040-3590.3.1.82>
- Manly, J. y Echemendia, R. (2007). Race-specific norms: Using the model of hypertension to understand issues of race, culture, and education in neuropsychology. *Archives of Clinical Neuropsychology*, 22(3), 319-325.
<https://doi.org/10.1016/j.acn.2007.01.006>
- Martínez Arias, M. R., Hernández Lloreda, M. J. y Hernández Lloreda, M. V. (2006). *Psicometría*. Alianza Editorial.
- Matthews, C. G. y Klove, H. (1964). Instruction manual for the adult neuropsychology test battery. *Madison, WI: University of Wisconsin Medical School*, 36.
- McCaffrey, R. J., Duff, K., & Westervelt, H. J. (Eds.). (2000). *Practitioner's Guide to Evaluating Change with Neuropsychological Assessment Instruments*. Springer US. <https://doi.org/10.1007/978-1-4615-4233-9>
- Mccaffrey, R. J., Ortega, A., Orsillo, S. M., Nelles, W. B. y Haase, R. F. (1992). Practice Effects in Repeated Neuropsychological Assessments. *The Clinical Neuropsychologist*, 6(1), 32-42. <https://doi.org/10.1080/13854049208404115>

- McCaffrey, R. J. y Westervelt, H. J. (1995). Issues associated with repeated neuropsychological assessments. *Neuropsychology Review*, 5(3), 203-221. <https://doi.org/10.1007/BF02214762>
- McSweeny, A. J., Naugle, R. I., Chelune, G. J. y Lüders, H. (1993). «T Scores for Change»: An illustration of a regression approach to depicting change in clinical neuropsychology. *Clinical Neuropsychologist*, 7(3), 300-312. <https://doi.org/10.1080/13854049308401901>
- Miotto, E. C., Campanholo, K. R., Rodrigues, M. M., Serrao, V. T., Lucia, M. C. S. de y Scaff, M. (2012). Hopkins verbal learning test-revised and brief visuospatial memory test-revised: preliminary normative data for the Brazilian population. *Arquivos de Neuro-Psiquiatria*, 70(12), 962-965. <https://doi.org/10.1590/S0004-282X2012001200014>
- Mitrushina, M., Boone, K. B., Razani, J. y D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment (2nd edition)*. Oxford Uni.
- Moering, R. G., Schinka, J. A., Mortimer, J. A. y Graves, A. B. (2004). Normative data for elderly African Americans for the Stroop Color and Word Test. *Archives of Clinical Neuropsychology*, 19(1), 61-71. [https://doi.org/10.1016/S0887-6177\(02\)00219-6](https://doi.org/10.1016/S0887-6177(02)00219-6)
- Monaci, L. y Morris, R. G. (2012). Neuropsychological screening performance and the association with activities of daily living and instrumental activities of daily living in dementia: baseline and 18- to 24-month follow-up. *International Journal of Geriatric Psychiatry*, 27(2), 197-204. <https://doi.org/10.1002/gps.2709>
- Montoya-Murillo, G., Ibarretxe-Bilbao, N., Peña, J. y Ojeda, N. (2019). Effects of cognitive rehabilitation on cognition, apathy, quality of life, and subjective

complaints in the elderly: a randomized controlled trial. *The American journal of geriatric psychiatry*, 1-12. <https://doi.org/10.1016/j.jagp.2019.10.011>

Mungas, D., Marshall, S. C., Weldon, M., Haan, M. y Reed, B. R. (1996). Age and education correction of Mini-Mental State Examination for English- and Spanish-speaking elderly. *Neurology*, 46(3), 700-706.
<https://doi.org/10.1212/WNL.46.3.700>

Munro, C. A., Winicki, J. M., Schretlen, D. J., Gower, E. W., Turano, K. A., Muñoz, B., Keay, L., Bandeen-Roche, K. y West, S. K. (2012). Sex differences in cognition in healthy elderly individuals. *Aging, Neuropsychology, and Cognition*, 19(6), 759-768. <https://doi.org/10.1080/13825585.2012.690366>

Naglieri, J. A. y Das, J. P. (1997). *Cognitive assessment system: Interpretive handbook*. Riverside.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L. y Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, 53(4), 695-699. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>

Neil-Pirozzi, T. M. O., Goldstein, R., Strangman, G. E. y Glenn, M. B. (2012). Test – re-test reliability of the Hopkins Verbal Learning Test-Revised in individuals with traumatic brain injury. *Brain Injury*, 26(12), 1425-1430.
<https://doi.org/10.3109/02699052.2012.694561>

Nieto, A., Hernández-Rodríguez, E., Hernández-Torres, A., Rodríguez-Solís, P. V., Hess-Medler, S., Machado-Fernández, A., Molina-Rodríguez, Y. y Barroso, J. (2014). Versión paralela del test de aprendizaje verbal España-Complutense

(TAVEC). *Revista de Neurologia*, 58(9), 389-395. <https://doi.org/rn2014432> [pii]

Norman, M. A., Moore, D. J., Taylor, M., Franklin, D., Cysique, L., Ake, C., Lazzarotto, D., Vaida, F. y Heaton, R. K. (2011). Demographically Corrected Norms for African Americans and Caucasians on the Hopkins Verbal Learning Test-Revised, Brief Visuospatial Memory Test-Revised, Stroop Color and Word Test, and Wisconsin Card Sorting Test 64-Card Version. *Journal of Clinical and Experimental Neuropsychology*, 33(7), 793-804.
<https://doi.org/10.1080/13803395.2011.559157>

Nuechterlein, K. H. y Green, M. F. (2006). *MATRICES Consensus Cognitive Battery*. MATRICS Assessment, Inc.

Nuechterlein, K. H., Green, M. F., Kern, R. S., Baade, L. E., Barch, D. M., Cohen, J. D., Essock, S., Fenton, W. S., Frese, F. J., Gold, J. M., Goldberg, T., Heaton, R. K., Keefe, R. S. E., Kraemer, H., Mesholam-Gately, R., Seidman, L. J., Stover, E., Weinberger, D. R., Young, A. S., ... Marder, S. R. (2008). The MATRICS Consensus Cognitive Battery, Part 1: Test Selection, Reliability, and Validity. *Am J Psychiatry*, 165(2), 203–213.

Ojeda, N., Aretouli, E., Peña, J. y Schretlen, D. J. (2016). Age differences in cognitive performance: A study of cultural differences in Historical Context. *Journal of Neuropsychology*, 10(1), 104-115. <https://doi.org/10.1111/jnp.12059>

Ojeda, N., del Pino, R., Ibarretxe-Bilbao, N., Schretlen, D. J. y Peña, J. (2016). Test de evaluación cognitiva de Montreal: normalización y estandarización de la prueba en población española. *Revista de Neurología*, 63(11), 488.
<https://doi.org/10.33588/rn.6311.2016241>

Ojeda, N., Del Pino, R. y Peña, J. (2014). *UD Interferencia*. (Patent N.º N°BI-895-13).

- Osterrieth, P. A. (1944). Le test de copie d'une figure complexe; contribution à l'étude de la perception et de la mémoire. *Archives de Psychologie*, 30, 206-356.
- Pachana, N. A., Thompson, L. W., Marcopulos, B. A. y Yoash-Gantz, R. (2004). California Older Adult Stroop Test (COAST). *Clinical Gerontologist*, 27(3), 3-22. https://doi.org/10.1300/J018v27n03_02
- Pakhomov, S. V. S., Eberly, L. y Knopman, D. (2016). Characterizing cognitive performance in a large longitudinal study of aging with computerized semantic indices of verbal fluency. *Neuropsychologia*, 89, 42-56. <https://doi.org/10.1016/j.neuropsychologia.2016.05.031>
- Parmenter, B. A., Testa, S. M., Schretlen, D. J., Weinstock-Guttman, B. y Benedict, R. H. B. (2010). The utility of regression-based norms in interpreting the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). *Journal of the International Neuropsychological Society*, 16(1), 6-16. <https://doi.org/10.1017/S1355617709990750>
- Peña-Casanova, J. (1991). Programa integrado de exploración neuropsicológica - test barcelona: bases teóricas, objetivos y contenidos. *Revista de Logopedia, Foniatria y Audiologia*, 11(2), 66-79. [https://doi.org/10.1016/S0214-4603\(91\)75505-8](https://doi.org/10.1016/S0214-4603(91)75505-8)
- Peña-Casanova, Jordi, Blesa, R., Aguilar, M., Gramunt-Fombuena, N., Gómez-Ansón, B., Oliva, R., Molinuevo, J. L., Robles, A., Barquero, M. S., Antúnez, C., Martínez-Parra, C., Frank-García, A., Fernández, M., Alfonso, V. y Sol, J. M. (2009). Spanish multicenter normative studies (NEURONORMA project): Methods and sample characteristics. *Archives of Clinical Neuropsychology*, 24(4), 307-319. <https://doi.org/10.1093/arclin/acp027>
- Peña-Casanova, Jordi, Casals-Col, M., Quintana, M., Sánchez-Benavides, G., Rognoni,

- T., Calvo, L., Palomo, R., Aranciva, F., Tamayo, F. y Manero, R. M. (2012). Estudios normativos españoles en población adulta joven (Proyecto NEURONORMA jóvenes): métodos y características de la muestra. *Neurología*, 27(5), 253-60. <https://doi.org/10.1016/j.nrl.2011.12.019>
- Peña-Casanova, Jordi, Quiñones-Úbeda, S., Gramunt-Fombuena, N., Quintana, M., Aguilar, M., Molinuevo, J. L., Serradell, M., Robles, A., Barquero, M. S., Payno, M., Antúnez, C., Martínez-Parra, C., Frank-García, A., Fernández, M., Alfonso, V., Sol, J. M. y Blesa, R. (2009). Spanish Multicenter Normative Studies (NEURONORMA Project): Norms for the Stroop Color-Word Interference Test and the Tower of London-Drexel. *Archives of Clinical Neuropsychology*, 24(4), 413-429. <https://doi.org/10.1093/arclin/acp043>
- Peña, J., del Pino, R., Ibarretxe-Bilbao, N., Schretlen, D. J. y Ojeda, N. (2016). Test de comparación perceptual de Salthouse: normalización y estandarización de la prueba en población española. *Revista de Neurología*, 62(01), 13. <https://doi.org/10.33588/rn.6201.2015307>
- Portellano, J. A. (2005). *Introducción a la neuropsicología*. McGraw-Hill España.
- Reilly, D., Neumann, D. L. y Andrews, G. (2017). Gender Differences in Spatial Ability: Implications for STEM Education and Approaches to Reducing the Gender Gap for Parents and Educators. En M. Khine (Ed.), *Visual-spatial Ability in STEM Education* (pp. 195-224). Springer International Publishing. https://doi.org/10.1007/978-3-319-44385-0_10
- Reitan, R. M. (1985). *The Halstead-Reitan Neuropsychological Test Battery*. Neuropsychology Press.
- Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique.

(Les problems). *Archives de Psychologie*, 28, 215-285.

Rivera, D., Morlett-Paredes, A., Peñalver Guía, A. I., Irías Escher, M. J., Soto-Añari, M., Aguayo Arellis, A., Rute-Pérez, S., Rodríguez-Lorenzana, A., Rodríguez-Agudelo, Y., Albaladejo-Blázquez, N., García De La Cadena, C., Ibáñez-Alfonso, J. A., Rodríguez-Irizarry, W., García-Guerrero, C. E., Delgado-Mejía, I. D., Padilla-López, A., Vergara-Moragues, E., Barrios Nevado, M. D., Saracosti Schwartzman, M. y Arango-Lasprilla, J. C. (2017). Stroop Color-Word Interference Test: Normative data for Spanish-speaking pediatric population. *NeuroRehabilitation*, 41(3), 605-616. <https://doi.org/10.3233/NRE-172246>

Rivera, D., Perrin, P. B., Stevens, L. F., Garza, M. T., Weil, C., Saracho, C. P., Rodríguez, W., Rodríguez-Agudelo, Y., Rábago, B., Weiler, G., García De La Cadena, C., Longoni, M., Martínez, C., Ocampo-Barba, N., Aliaga, A., Galarza-Del-Angel, J., Guerra, A., Esenarro, L. y Arango-Lasprilla, J. C. (2015). Stroop Color-Word Interference Test: Normative data for the Latin American Spanish speaking adult population. *NeuroRehabilitation*, 37(4), 591-624. <https://doi.org/10.3233/NRE-151281>

Rognoni, T., Casals-Coll, M., Sánchez-Benavides, G., Quintana, M., Manero, R. M., Calvo, L., Palomo, R., Aranciva, F., Tamayo, F. y Peña-Casanova, J. (2013). Estudios normativos españoles en población adulta joven (proyecto NEURONORMA jóvenes): normas para las pruebas Stroop Color-Word Interference Test y Tower of London-Drexel University. *Neurología*, 28(2), 73-80. <https://doi.org/10.1016/j.nrl.2012.02.009>

Rönnlund, M. y Nilsson, L. G. (2006). Adult life-span patterns in WAIS-R Block Design performance: Cross-sectional versus longitudinal age gradients and relations to demographic factors. *Intelligence*, 34(1), 63-78.

<https://doi.org/10.1016/j.intell.2005.06.004>

- Ryan, J., Woods, R. L., Murray, A. M., Shah, R. C., Britt, C. J., Reid, C. M., Wolfe, R., Nelson, M. R., Lockery, J. E., Orchard, G., Trevaks, R. E., Chong, T. J., Mcneil, J. J., Storey, E. y ASPREE Investigator Group. (2020). Normative performance of older individuals on the Hopkins Verbal Learning Test-Revised (HVLTR) according to ethno-racial group, gender, age and education level. *The Clinical Neuropsychologist*, 0(0), 1-17. <https://doi.org/10.1080/13854046.2020.1730444>
- Sachs, B. C., Lucas, J. A., Smith, G. E., Ivnik, R. J., Petersen, R. C., Graff-Radford, N. R. y Pedraza, O. (2012). Reliable Change on the Boston Naming Test. *Journal of the International Neuropsychological Society*, 18(2), 375-378. <https://doi.org/10.1017/S1355617711001810>
- Salthouse, T. A. (2010). Influence of age on practice effects in longitudinal neurocognitive change. *Neuropsychology*, 24(5), 563-572. <https://doi.org/10.1037/a0019026>
- Salthouse, T. A. (2019). Trajectories of normal cognitive aging. *Psychology and Aging*, 34(1), 17-24. <https://doi.org/10.1037/pag0000288>
- Salthouse, T. A. y Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, 27(5), 763-776. <https://doi.org/10.1037/0012-1649.27.5.763>
- Sánchez-Benavides, G., Peña-Casanova, J., Casals-Coll, M., Gramunt, N., Manero, R. M., Puig-Pijoan, A., Aguilar, M., Robles, A., Antúnez, C., Frank-García, A., Fernández-Martínez, M. y Blesa, R. (2016). One-Year Reference Norms of Cognitive Change in Spanish Old Adults : Data from the NEURONORMA Sample. *Archives of Clinical Neuropsychology*, 31(4), 378–388.

<https://doi.org/10.1093/arclin/acw018>

Schretlen, D. J. (2010). *Modified Wisconsin Card Sorting Test (M-WCST): professional manual*. Psychological Assessment Resources Inc.

Schretlen, D. J., Bobholz, J. H. y Brandt, J. (1996). Development and psychometric properties of the brief test of attention. *The Clinical Neuropsychologist*, 10(1), 80-89. <https://doi.org/10.1080/13854049608406666>

Schretlen, D. J., Cascella, N. G., Meyer, S. M., Kingery, L. R., Testa, S. M., Munro, C. A., Pulver, A. E., Rivkin, P., Rao, V. A., Diaz-Asper, C. M., Dickerson, F. B., Yolken, R. H. y Pearlson, G. D. (2007). Neuropsychological Functioning in Bipolar Disorder and Schizophrenia. *Biological Psychiatry*, 62(2), 179-186. <https://doi.org/10.1016/j.biopsych.2006.09.025>

Schretlen, D. J., Testa, M. y Pearlson, G. D. (2010). *Calibrated Neuropsychological Normative System professional manual*. PAR.

Schretlen, D. J. y Vannorsdall, T. D. (2010). Calibrated ideational fluency assessment (CIFA): professional manual. *Lutz, FL: Psychological Assessment Resources*.

Sisco, S. M., Slonena, E., Okun, M. S., Bowers, D. y Price, C. C. (2016). Parkinson's Disease and the Stroop Color Word Test: Processing Speed and Interference Algorithms. *The Clinical Neuropsychologist*, 30(7), 1104-1117. <https://doi.org/10.1080/13854046.2016.1188989>

Smerbeck, A. M., Parrish, J., Yeh, E. A., Weinstock-Guttman, B., Hoogs, M., Serafin, D., Krupp, L. y Benedict, R. H. B. (2012). Regression-Based Norms Improve the Sensitivity of the National MS Society Consensus Neuropsychological Battery for Pediatric Multiple Sclerosis (NBPMS). *The Clinical Neuropsychologist*, 26(6), 985–1002. <https://doi.org/10.1080/13854046.2012.704074>

- Steinberg, B. A., Bieliauskas, L. A., Smith, G. E. y Ivnik, R. J. (2005). Mayo's Older Americans Normative Studies: Age- and IQ-adjusted norms for the Trail-Making Test, the Stroop Test, and MAE Controlled Oral Word Association Test. *Clinical Neuropsychologist*, 19(3-4), 329-377. <https://doi.org/10.1080/13854040590945210>
- Strauss, E., Sherman, E. M. y Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary (3rd edition)*. Oxford Uni.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643-662. <https://doi.org/10.1037/h0054651>
- Taylor, L. B. (1969). Localisation of cerebral lesions by psychological testing. *Clinical neurosurgery*, 16, 269-287.
- Temkin, N. R., Heaton, R. K., Grant, I. y Dikmen, S. S. (1999). Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neuropsychological Society*, 5(4), 357-369. <https://doi.org/10.1017/S1355617799544068>
- Testa, S. M., Winicki, J. M., Pearlson, G. D., Gordon, B. y Schretlen, D. J. (2009). Accounting for estimated IQ in neuropsychological test performance with regression-based techniques. *Journal of the International Neuropsychological Society*, 15(6), 1012 – 1022. <https://doi.org/10.1017/S1355617709990713>
- Troyer, A. K., Leach, L. y Strauss, E. (2006). Aging and response inhibition: Normative data for the Victoria Stroop Test. *Aging, Neuropsychology, and Cognition*, 13(1), 20-35. <https://doi.org/10.1080/138255890968187>
- Urbina, S. (2004). *Essentials of psychological testing*. John Wiley & Sons.
- van Boxtel, M. P. J., ten Tusscher, M. P. M., Metsemakers, J. F. M., Willems, B. y Jolles, J. (2001). Visual Determinants of Reduced Performance on the Stroop

- Color-Word Test in Normal Aging Individuals. *Journal of Clinical and Experimental Neuropsychology*, 23(5), 620-627.
<https://doi.org/10.1076/jcen.23.5.620.1245>
- van de Vijver, F. y Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *European Review of Applied Psychology*, 54(2), 119-135. <https://doi.org/10.1016/j.erap.2003.12.004>
- Van der Elst, W., Van Boxtel, M. P. J., Van Breukelen, G. J. P. y Jolles, J. (2006). The Stroop Color-Word Test. *Assessment*, 13(1), 62-79.
<https://doi.org/10.1177/1073191105283427>
- Vanderploeg, R. D., Schinka, J. A., Jones, T., Small, B. J., Graves, A. B. y Mortimer, J. A. (2000). Elderly Norms for the Hopkins Verbal Learning Test-Revised. *The Clinical Neuropsychologist*, 14(3), 318-324. [https://doi.org/10.1076/1385-4046\(200008\)14](https://doi.org/10.1076/1385-4046(200008)14)
- Villa Rodríguez, M. Á., Navarro Calvillo, M. E. y Villaseñor Cabrera, T. de J. (2017). *Neuropsicología clínica hospitalaria*. El Manual Moderno.
- Waldrop-Valverde, D., Ownby, R. L., Jones, D. L., Sharma, S., Nehra, R., Kumar, A. M., Prabhakar, S. y Kumar, M. (2015). Neuropsychological test performance among healthy persons in northern India : development of normative data. *J. Neurovirol*, 21(4), 433-438. <https://doi.org/10.1007/s13365-015-0332-4>
- Wasserman, V., Emrani, S., Matusz, E. F., Miller, D., Garrett, K. D., Gifford, K. A., Hohman, T. J., Jefferson, A. L., Au, R., Swenson, R. y Libon, D. J. (2019). Visual and Verbal Serial List Learning in Patients with Statistically-Determined Mild Cognitive Impairment. *Innovation in Aging*, 3(2), 1-12.
<https://doi.org/10.1093/geroni/igz009>

- Watkins, M. W. y Smith, L. G. (2013). Long-term stability of the Wechsler Intelligence Scale for Children - Third Edition. *Psychological Assessment*, 25(2), 477-483.
<https://doi.org/10.1037/a0031653>
- Wechsler, D. (2009). *Wechsler Memory Scale-Fourth Edition (WMS-IV)*. Pearson Assessment.
- Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W. y Delazer, M. (2003). Sex differences in cognitive functions. *Personality and Individual Differences*, 35(4), 863-875. [https://doi.org/10.1016/S0191-8869\(02\)00288-X](https://doi.org/10.1016/S0191-8869(02)00288-X)
- Zachary, R. A. y Gorsuch, R. L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology*, 41(1), 86-94.
[https://doi.org/10.1002/1097-4679\(198501\)41:1<86::AID-JCLP2270410115>3.0.CO;2-W](https://doi.org/10.1002/1097-4679(198501)41:1<86::AID-JCLP2270410115>3.0.CO;2-W)
- Zec, R. F. (1995). The neuropsychology of aging. *Experimental Gerontology*, 30(3-4), 431-442. [https://doi.org/10.1016/0531-5565\(94\)00066-C](https://doi.org/10.1016/0531-5565(94)00066-C)
- Zec, R. F., Markwell, S. J., Burkett, N. R. y Larsen, D. L. (2005). A longitudinal study of confrontation naming in the “normal” elderly. *Journal of the International Neuropsychological Society*, 11(06). <https://doi.org/10.1017/S1355617705050897>
- Zgaljardic, D. J. y Benedict, R. H. B. (2001). Evaluation of Practice Effects in Language and Spatial Processing Test Performance. *Applied Neuropsychology*, 8(4), 218-223. https://doi.org/10.1207/S15324826AN0804_4
- Zucchella, C., Capone, A., Codella, V., De Nunzio, A. M., Vecchione, C., Sandrini, G., Pace, A., Pierelli, F. y Bartolo, M. (2013). Cognitive rehabilitation for early post-surgery inpatients affected by primary brain tumor: a randomized, controlled trial. *Journal of Neuro-Oncology*, 114(1), 93-100. <https://doi.org/10.1007/s11060-013->

1153-z

Zucchella, C., Federico, A., Martini, A., Tinazzi, M., Bartolo, M. y Tamburin, S.

(2018). Neuropsychological testing. *Practical Neurology*, 18(3), 227-237.

<https://doi.org/10.1136/practneurol-2017-001743>

Anexo I



Instrucciones para la administración del test UD Interferencia.

Los ítems deben ser leídos por el examinado en columnas y no se le debe permitir guiarse por el dedo. El tiempo de administración de cada lámina es de 30 segundos y se registra el número de ítems leídos/denominados correctamente en ese tiempo, que componen la puntuación de cada lámina. El orden de administración de las láminas y las instrucciones que se le proporcionan al participante son las siguientes:

Lámina 1 - Palabras (P):

“En esta parte, le voy a pedir que lea, a la mayor velocidad que pueda y en voz alta, estas palabras. Va a empezar leyendo desde aquí, de una en una hasta llegar al final de la columna. Cuando llegue al final, continuará por el principio hasta el final y en el mismo orden hasta que yo diga PARE. Entonces me tendrá que señalar con el dedo donde se ha quedado. Si llega al final (señala el final de la última columna) y no le he dicho que pare, vuelva a empezar por el principio y continúe en el mismo orden”.

- No permitirle guiarse por el dedo.
- En caso de saltarse algún color, contarlos y después descontarlos.
- Tiempo límite: 30 segundos.

Examinador: anote en las hojas de respuesta el nº total de palabras leídas por el sujeto.



Lámina 2 - Colores (C):

“En esta parte, quiero que lea lo más rápido que pueda y en voz alta, el color de la tinta de los círculos, de forma que va a empezar aquí diciéndome ROSA, AZUL, NEGRO... y así de una en una hasta llegar al final de la columna. Cuando llegue al final, continuará por el principio hasta el final y en el mismo orden hasta que yo le diga PARE. En ese momento, me tendrá que señalar con el dedo dónde se ha quedado”.

- No permitirle guiarse por el dedo.
- En caso de saltarse algún color, contarlo y después descontarlo.
- Tiempo límite: 30 segundos.

Examinador: anote en las hojas de respuesta el nº total de colores leídos por el sujeto.

Lámina 3 - Palabras-colores (PC):

“En esta parte quiero que diga lo más rápido que pueda y en voz alta, el color de la tinta con la que están escritas estas palabras, de manera que en la primera me va a decir: ROSA, AZUL, NEGRO... Debe empezar con la primera palabra de una en una, siguiendo el orden de la columna hasta el final y cuando llegue a la última, continúa por la siguiente columna en el mismo orden. Cuando le diga PARE me va a señalar dónde se ha quedado. En esta parte es muy fácil equivocarse. Si se equivoca, yo le señalaré la palabra en la que se ha equivocado, lo antes que pueda, me dice de nuevo el color de la tinta de esa palabra y continua en el mismo orden. ¿Lo ha entendido?”.

- No permitirle guiarse por el dedo.
- En caso de equivocarse en algún color, señalarle la palabra para que la repita.
- Tiempo límite: 30 segundos.

Examinador: anote en las hojas de respuesta el nº total de respuestas correctas dadas por el sujeto.