Heliyon 9 (2023) e17435

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Identification and comparison of the main variables affecting early university dropout rates according to knowledge area and institution

Alba Llauró [a,*], David Fonseca [a], Susana Romero [b], Marian Aláez [b], Jorge Torres Lucas [c], María Martínez Felipe [c]

[a] *Technology Enhanced Learning Research Line of Human Environment Research Group (HER-TEL), La Salle, Ramon Llull University, Barcelona, Spain*

[b] *Unidad de Innovación Docente, Deusto University, Bilbao, Spain*

[c] *Grupo de Investigación: Innovación Educativa para Una Formación Integral en Diferentes Etapas y Contextos. Departamento de Educación y Métodos de Investigación y Evaluación Universidad Pontificia Comillas, Madrid, Spain*

A R T I C L E I N F O

Keywords:
Academic analytics
Early dropout
First-year students
Learning analytics
Predictions
Tutoring

A B S T R A C T

The dropout rate in universities is a widely studied issue that concerns both universities and public organizations. Most studies focus on quantifying the phenomenon and identifying the variables involved. This paper uses a multidisciplinary approach to parameterize the factors that define the entry profile of undergraduates at the national level in Spain in collaboration with three universities in different regions and with different disciplines. The aim is to reduce the dropout rate in the first year of study towards a degree. The research questions focus on the weighting of personal variables about students by tutors and whether there are differences in the weighting systems for the main variables as differentiated by discipline, university and/or region. The document is organized to describe the method and context of the study, present the main results, show the application of the survey instrument in a case study, and provide conclusions. The method is based on the two fundamental factors, including the influence of certain student characteristics at matriculation and the importance of a positive experience in the first year of the degree. The study is focused on two elements that inspire the current proposal: the need to identify and rapidly detect students who, due to their entry characteristics, are at a greater risk of dropout and the importance of guaranteeing a good start in the first year of the degree. The study uses a multidisciplinary approach and combines qualitative and quantitative methods to gather data. The study also uses a survey instrument that measures the risk of student dropout based on the weighting of personal variables by tutors. The results of the study have allowed us to categorize the main variables of the student profile that affect the risk of dropout and establish them as aspects to be monitored by the tutors in the first weeks. Furthermore, it has been shown that although there are no significant differences in the averages of dropout risk calculated with either global or specific weighting systems (by centre or studies), there is a tendency observed by the tutors that the weighted averages generated by disciplinary focus are closer to identifying the student's real risk of dropping out.

* Corresponding author.
E-mail address: alba.llauro@salle.url.edu (A. Llauró).

https://doi.org/10.1016/j.heliyon.2023.e17435
Received 18 January 2023; Received in revised form 13 June 2023; Accepted 16 June 2023
Available online 17 June 2023

1. Introduction

Dropout in universities is a widely analysed issue that on the one hand, interests universities to incorporate measures that promote academic success rates and on the other hand, interests public organizations in many countries due to the professional, social, and economic repercussions associated with this possible dropout [1–3]. Identifying the causes of early dropout allows universities to focus their efforts on honing their level of excellence, which can lead to an improvement in their positioning in quality rankings and, therefore, in their prestige [4].

Currently, most studies focus mainly on trying to quantify the phenomenon of early dropout, identifying the variables involved, and constructing and validating explanatory models of this phenomenon [5,6]. The interest in this line of research is further accentuated by the COVID-19 pandemic, as educational systems had to urgently adapt to new needs, which often did not ensure good instructional design with the support of ICTs [7,8].

According to the data presented by the Spanish Ministry of Science, Innovation and Universities in 2020, a dropout rate of 33.9% was revealed for the cohort newly enrolled in the 2013-14 academic year. This percentage increased to 35% in the case of public universities and decreased to 27.5% in private universities. These percentages are higher than the 24% recorded in the average dropout rate collected by the Organization for Economic Cooperation and Development [9].

Briefly, early withdrawal from studies is a key issue for the university system, with very negative consequences at both the individual and institutional levels. It is a problem marked by its great complexity and with undeniable multicausal variables, as we can see in the previous studies. Based on the data obtained from previous reports and referenced studies, the motivation for this study increases when the scope of analysis focuses on private institutions (which thus have high financial dependence on student income) and technical studies (historical areas with high drop-out rates). It is of vital importance to generate an instrument that allows a reliable indicator of potential dropout to be obtained and that is not limited to a local setting or specific studies, as many of the previous benchmarks, but can be evaluated in parallel in different centres, regions and studies to generate a reliable and replicable instrument.

This paper focuses on multidisciplinary research to parameterize the factors that define the entry profile of undergraduate students at the Spanish national level. The main hypothesis is that it is possible to identify those personal variables in the student's profile that are initially indicators of a higher risk of early dropout, so that from this identification, academic tutors have information on which to act and potentially mitigate this risk. From this hypothesis, a secondary hypothesis is defined that is based on the idea that an adjusted prioritization of weights of these variables by the tutors is more accurate when applied to students in the same field of knowledge than the classification carried out by the tutors. Therefore, for this research, the following research questions (RQ) have been defined.

RQ1. In the construction of an instrument that measures the risk of student dropout based on the weighting of personal variables by tutors, which variables have the greatest weight?

RQ2. Are there significant differences in the weight of the main variables, based on criteria set by tutors for areas of knowledge, universities and/or regions, in such a way that it affects the result of the criticality calculation?

This document is organized as follows: Section II describes the method and context in which the study is carried out. Section III explains the main results. Section IV shows the application of the instrument to a case study. Finally, Section V presents the conclusions of the research.

2. Theoretical context

This research is at the intersection of three main fields of knowledge. Educational data mining is a central field, but this inquiry is focused specifically for the purpose of helping improve academic management (academic analytics) and therefore for the improvement of the student's curriculum (learning analytics). Additionally, we are at the leading edge of studies of psycho-pedagogical knowledge, such as tutoring and activities/services for curricular accompaniment that certain teachers use to accompany, advise and guide students in those academic decisions that may affect their performance and therefore their curricular success. Finally, a process that integrates data, monitors results, and helps decision-making based on a user-centred design (UCD), which is based on both qualitative and quantitative mixed processes, is needed to achieve satisfactory and efficient user experiences. We will now briefly contextualize each of the three areas described.

As we have introduced, our research works with personal variables and academic data to improve the educational process, a term that is circumscribed by the theoretical framework of academic analytics. In this sense, the processes linked to academic analytics are defined as those that evaluate and interpret all types of data to improve decision-making and ultimately academic processes beyond student learning, which would fall within the scope of learning analytics [10,11].

Academic analytics is a hybrid approach that provides data to higher education institutions to improve operational and financial decision-making. While learning analytics is more concerned with course- and department-level data [12–14], academic analytics is more concerned with student profiles, academic performance, and knowledge flow. Academic analytics aims to analyse data from student interactions to improve educational, academic, and teaching-related processes [15]. The management of these data provides critical information to educational institutions to make decisions to improve programs and student-tracking to maximize student achievement [16]. In both research and practice, learning and academic analytics has proven useful in identifying variables that influence learning outcomes and establish relationships between competencies, educational methodologies, and curricular structures. These analyses provide information to personalize courses and detect at-risk students to provide early intervention. In this way, it is also possible to improve teaching to retain more students throughout the course [14,17].

The second main area of study is related to tutoring services. This activity has been gaining importance in recent years and has been particularly relevant during the period of the COVID-19 pandemic [18], The student's motivation, their mood resulting from the lockdown period, the new methods of work and use for the materials, and other studied factors are reasons why it is now more important than ever to help students by providing tutoring services [19,20]. This service is considered a very important intervention in student learning throughout their studies, especially at preuniversity levels, but it is gaining importance in recent years in university courses, in which there are similar needs to detect problems and monitor students, but with clearly differentiated processes [21].

The need for generic content in preuniversity studies is recognized by students; however, in a significant number of cases, students do not find meaning in their choice of university degree, especially when they have chosen a degree with technical-scientific-technological content. This fact, combined with the difficulty associated with the educational level, leads to frustration. If other factors are added, such as incorrect or insufficiently adapted study habits, lack of knowledge of how to cope with occasional failures, distance from the family environment, greater freedom of movement, etc. [22], the result is students' lack of adaptation to university challenges.

Therefore, tutoring in the first year of university is of particular importance. The tutor can advise the student on the most critical points of the course, as well as personalize the activity to generate a greater impact and help the student overcome the first year with less difficulty [23]. If the tutor can collect, analyse, and manage data related to their students' admission profile, they can anticipate necessary actions during the course for those students who may be at risk of dropping out or affected by a situation that may increase this risk.

Finally, the third main knowledge field is related to user experience as a framework. User experience is a discipline that considers people's perceptions and responses to interactive behaviour with a service, design, or proposal [24]. In this sense, usability studies consider both factors related to the process and those related to users' emotions and perceptions. The method applied in the study is based on an iterative and participatory design, where the selected variables provide detailed information about the student's profile. Based on this premise, the user-centred design (UCD) methodology [25] is a philosophy that considers the user as part of the service creation process, providing their motivations, needs, or desires during each phase.

As we have introduced, among the possible methodological designs for monitoring students, the iterative design on which the proposal is based stands out as one of the most practical, as it allows for greater periodic data collection and validation by repetition and consolidation of data [26]. On the other hand, participatory design actively considers all involved parties [27]. Combining iterative and participatory approaches improves the data collection of any user-centred study, which, in the context of our research, is focused on the student as the user. The variables, indicators, and study values that are used have been defined in conjunction with tutors, students, and previous work so that the proposal refines methods to obtain the student's initial profile, allowing more efficient interventions to improve their performance and help them with the initial adaptation. In this sense, the critical review of previous research has been covered previously by the authors, published in Ref. [28].

The approach, which is based on these three areas of knowledge, is what clearly differentiates the proposal from the rest of the studies of university dropout rates. Additionally, the study works specifically with three of the main private universities in Spain, covering different regional data and educational fields: La Salle, Ramon Llull University (La Salle-RLU), Pontifical Comillas University (Comillas), and the University of Deusto.

La Salle-RLU, founded in 1965 by the Brothers of La Salle, offers undergraduate, graduate, postgraduate, doctoral, and specialization courses in the fields of Engineering, ICT and Computer Science, Architecture and Building, Business and Management, Animation & VFX, Digital Arts, Health Engineering, and Philosophy, with technology and humanism as essential elements of its DNA. The university welcomes 5109 students, both locals, representing 68.5% of the student body (87% from Catalonia and 13% from the rest of Spain), and international students from multiple nations, representing the other 31.5%. Three percent of students carry out their internships at the university.

Comillas University in Madrid, founded in 1890, is a Catholic university governed by the Society of Jesus. It offers undergraduate, postgraduate, and doctoral studies to 12,800 students in the following areas: Business, Social-Humanistic, Legal, Health, Engineering, and Theology. Twenty percent of its student body is international or participates in an international exchange program at the university.

Deusto University was founded in 1886 by the Society of Jesus. With campuses in Bilbao and San Sebastian and branches in Vitoria and Madrid, its hallmarks are education in competencies and values, thanks to a socially recognized pedagogical model. It offers various undergraduate, graduate, and doctoral programs to approximately 11,000 students, both local and international, distributed across 8 faculties and an affiliated centre: BAM (Begoñako Andramari), Health Sciences, Economics and Business Administration, Social and Human Sciences, Law, Education and Sports, Engineering, and Theology.

The three participating universities are part of the Aristos Campus Mundus (ACM) consortium, which has been accredited as an International Campus of Excellence in the regional European category (BOE-A-2015-13,413). Additionally, ACM has signed an agreement of advanced strategic cooperation with the North American universities Georgetown, Boston College and Fordham. The three universities in the study are private Spanish universities and therefore face similar issues. Private universities often have high enrolment fees, which may make them less accessible to students from low-income families. Additionally, financial aid is limited, and private universities may not have as much financial assistance available to students as public universities, which can also make them less accessible. Private universities may have a less diverse student body compared to public universities, as they may attract students from similar socioeconomic backgrounds. Private universities depend heavily on student enrolment for revenue, which leads to a high level of tracking of enrolled students. For these reasons, and due to the concern for retaining those students with potential for good academic performance throughout their university degree, good tutorial action is needed. The tutor is the person in charge of academic monitoring and mentoring the student, who may sometimes feel alone or unmotivated. These emotional factors may lead to student dropout, but it has been shown that with good tutorial action, the student has a greater chance of achieving their goal [29].

As we have introduced, there are two key issues in the different studies developed on the existence, relationship, and interaction of various factors maintaining the percentages of early university dropout: a) the influence of certain student characteristics at university entry on the greater risk of dropout; b) the importance of a good start in the first year of the degree, with a positive experience of the transition and entry into the university world, in the probability of staying in the university [30]. These two aspects make us focus on two fundamental elements that inspire the current proposal.

- First, there is a need to identify and rapidly detect students who, due to their entry characteristics, might be at greater risk of early dropout. This requires the development of instruments that allow for early detection as well as the establishment of structures for obtaining, managing, and using the data for analysis and necessary decision-making.
- Second, it is necessary to develop corrective guidance and interventions to prevent dropout and support students in their successful incorporation into university life. From this perspective, the literature covers approaches to manage the first year of the university student in a specific academic program, which are very suggestive. For example, EPAU ("First Year Experience in University") or FYE ("First Year Experience") are programs used in the English tradition.

In this sense, tutorial action appears to be one of the best instruments to act on the needs of students in their entry and transition to the university, allowing tutors to identify, prevent and correct the influence of possible risk factors that may cause early dropout [31]. Experiences in this field allow us to say that actively incorporating first-year academic tutors in the ordering, scoring, and establishment of relative weights of the different risk factors for possible early dropout is a coherent and valid approach.

3. Method

a) Work methodology

The method used in this study is based on an iterative and participatory design, where the variables selected provide detailed information about the student's profile. Starting from this premise, the user-centred design (UCD) methodology considers users as part of the service creation process and establishes their motivations, needs, or desires at each phase.

As the first phase of the UCD, there was a search, analysis, and selection of diverse variables used to profile students. The variables were selected through a modified Delphi procedure using expert users to determine the questionnaire content efficacy. The expert group creating the initial approach was selected based on years of experience in teaching university students and assessments by mentored students. In all cases, a minimum of 10 years and a minimum rating of 4.2 out of 5 were required. Twelve university education professionals from La Salle-RLU attended, including five tutors from the School of Engineering, four tutors from the Business School, two from the School of Architecture, and one tutor from the School of Digital Arts. According to authors such as Landeta, the number of tutors was considered sufficient. This method was chosen because its effectiveness has been widely demonstrated in educational research [32] and is based on the knowledge and consensus of the consulted group to make the study more reliable. It was concluded that this is a complex issue in which a large number of factors can intervene in different ways depending on the context [16].

As a first step, an initial questionnaire was created with 13 items derived from a review of literature on survey variables, grouped by their three dimensions of personal data, study habits, and motivation [33].

In the initial iteration, the groups were tasked with conducting a qualitative evaluation of the items. The primary researcher received the reviews via electronic mail. This iteration achieved 58% accuracy, as detailed in the 2021 published article [33]. Following the first iteration, a lack of understanding between the tutors and students was identified, prompting the creation of additional questions. This allowed for the creation of a more detailed student profile, as well as a more accurate prediction. Based on the feedback from experts, the questionnaire was redesigned to measure study variables and underwent a second validation. After two rounds, the data were statistically processed and returned to the experts to achieve optimal weighting. Based on the results obtained, a third questionnaire was created, incorporating personal questions without weights to provide tutors with more detailed information about different students, thereby enabling them to provide more specific assistance. The new questions and blocks were determined through a literature review and expert analysis, as outlined in the related article published in 2023 [28].

From the responses received after the second round and subsequent analysis, a questionnaire was developed for the second phase of the UCD. The 13 factors were grouped into their three main blocks: personal data, study habits, and personal motivations.

Although treated as three dimensions, the variables were related: for example, age and previous study background influence motivation, and study habits are related to gender [6]. The university stage is more difficult than previous stages, and not all students who arrive at the university have developed the necessary study habits and time management to overcome these new challenges. Some papers show the relationship between study habits, academic performance, and other variables, such as the entrance grade to the university or the country of origin, which seemed to influence student success differently depending on the field of knowledge [34]. For the motivation dimension, factors such as the choice of degree and vocation are some of the most important aspects when deciding to continue or abandon a degree when there are difficulties. This block captures important resilience factors that help avoid abandonment [35].

The conviction of student towards the chosen degree is one of those factors that deserves to be studied because there are studies that show opposing conclusions about its influence. Therefore, in Ref. [36], the authors conclude that it does not have a significant influence on academic performance. However, as the same authors point out, this conclusion opposes the one reached earlier by Ref. [37]. The authors justify their findings because when students reach university, they often find that reality does not match what they had imagined. Another factor such as the distance to the university, although not indicative of abandonment, has been shown to have a statistically significant negative influence [38]. Similarly, studies reveal a positive influence between the perception of a scholarship and academic success when a high level of grades is required to receive it [39].

In the new iteration that started in the current academic year, 2022–2023, we have started redefining the questionnaire to make it more accurate and consider new variables. Some of these new variables, which are presented in other studies, were not considered in our first pilot study [40,41]. These previous works show that most automatic predictive systems take into account very few variables, so instead of collecting only those that coincided in all the cases studied, we have opted for the sum of all of them. Other studies, such as [42], analysed domains of success such as academic or group well-being instead of dropout rates, so nonsuccess or nonwell-being have been considered contrastive indicators that something might go wrong in the student's environment.

Table 1
Student's questionnaire and average obtained for each variable and dimension considering the general prioritization of all tutors and the weight obtained. Grey variables are without weight in the calculation for the general dimension.

#	Questions	
[1.318 final weight]	**Personal Data (5.67 – Tutor average)**	
1	[0.805] Year of birth (1.83)	
2	[0.512] Gender (1.17)	
[1.430 final weight]	**University entrance (6.15 – Tutor average)**	
3	[0.225] How did you enter university? (3.33)	[2.85] 4. From which baccalaureate/Voc. Training/other degree? (0.193) From which university? In what way?
	In which language model did you take your baccalaureate or equivalent?	In which languages?
	If you come from another degree, please answer the following question:	What degree was it? Up to which full year did you complete?
5	[0.231] Do you consider yourself to be a student of (good, pass, outstanding, excellent …) (3.43)	
6	[0.348] How do you usually study (books, notes, classmates, etc.)? (5.15)	
7	[0.231] Average mark of university access (3.43)	
8	[0.201] In which country did you study the baccalaureate or the compulsory course? (2.98)	
	In which school did you study your last studies?	
[1.641 final weight]	**Current data (7.06 – Tutor average)**	
9	[0.213] What grade are you attending? (7.31)	Name of tutor
	Where do you usually live (family address)?	
10	[0.183] Where do you live during the course? (6.28)	
11	[0.174] How long does it take you to get to university? (5.98)	
12	[0.129] Do you have a scholarship? (4.43)	What type of scholarship?
13	[0.133] How would you describe your level of knowledge of basic computer tools? (4.57)	
14	[0.157] Do you have a computer at home? (5.39)	
15	[0.111] Do you have siblings or close friends who are or have been students on your course? (3.81)	
16	[0.193] What is your relationship with classmates like? (6.61)	
17	[0.145] Do you have anyone among your classmates whom you could ask for notes when you cannot go to the classroom? (4.96)	
18	[0.203] Do you feel integrated in the classroom group? (6.96)	
[1.809 final weight]	**Degree choice (7.78 – Tutor average)**	
19	[0.629] How confident/assured are you about your chosen degree? (2.11)	
20	[0.585] Were the studies you are pursuing listed as your first choice? (1.96)	What was your first choice?
	Why did you choose the degree you are currently studying?	
21	[0.596] The syllabus of the 1st that I am studying seems to me to … (2.00)	
[1.921 final weight]	**Study habits (8.26 – Tutor average)**	
22	[0.510] Do you complete the tasks given to you? (4.11)	
23	[0.517] Do you study and revise every day? (4.17)	
24	[0.276] Where do you usually study? (2.22)	Where?
	With whom do you usually study?	
25	[0.211] Do you expect to join an academy? (1.7)	
26	[0.407] How many days before an exam do you study? (3.28)	
[1.882 final weight]	**Time spent studying (8.09 – Tutor average)**	
27	[0.227] Do you have a job, paid or unpaid? (4.48)	What work? Do you think it will prevent you from following the course normally?
28	[*0.142*] Do you carry out any activity to which you dedicate a significant amount of time, and which may take time away from your university studies? *(*2.81*)*	Please indicate which one(s) you carry out: How many hours/weeks do you spend on these activities?
29	[*0.281*] Are there any other circumstances (family, illness, etc.) that might make it difficult for you to follow your studies? *(*5.56*)*	
30	[*0.276*] How many hours do you plan to dedicate to study during the week? *(*5.46*)*	
31	[*0.209*] Are there any classes that you already know you will not be able to attend? *(*4.13*)*	
32	[*0.333*] Are you motivated to study sufficiently for the courses you have enrolled in? *(*6.57*)*	
33	[*0.209*] What are your goals in your chosen studies? *(*4.13*)*	
34	[*0.205*] Were you aware of how much time your studies might require? *(*4.06*)*	

After the first round of the project with 13 items and the detection of more factors in the influence of early dropout from university degrees, a new expert analysis was carried out qualitatively and various related studies were reviewed to determine new items to be considered in the new questionnaire. This expert analysis was carried out by 14 tutors (2 for each area of study included in the project) with 10 years of experience and a student rating of at least 4.2 out of 5, who have accumulated experience on the different profiles of incoming students from Deusto University, Comillas University and La Salle-RLU. Tutors are selected due to their proximity to students compared to other teachers, as they are the closest figure to the student and responsible for monitoring student progress, making them knowledgeable of situations where the student may potentially drop out of the program. The possible items were selected through a literature review, and thanks to the focus group with the experts, they were accepted or discarded.

Using the collected variables, a new questionnaire consisting of six dimensions was developed: personal data, university access, current data, degree choice, study habits, and time dedication. Three new dimensions were created, and new questions were added to each of the dimensions. The personal section of the initial iteration was split into two dimensions, personal and university access, to differentiate between intrinsic variables and those that have been developed during past courses. These dimensions provide a past image of the student, allowing us to understand how they have fared in previous courses and how prepared they are to undertake the selected degree. A new dimension, current data, was added, where questions related to the student's current personal situation are asked of the student, allowing the tutor a deeper understanding. For the motivation dimension, the name was changed to career choice, and new questions were added, in addition to redistributing some items to other dimensions. The study habits dimension remained, but new questions were added to obtain more information about the student. Finally, we have the last dimension, time dedicated to study, which provides information about the time the student can or intends to invest in their studies. These dimensions can be observed in the TEEM 2023 article [28].

All of these dimensions provide a clear picture of the student and allow for a more accurate assessment of the probability of dropout, as we have much more data available. These dimensions provide information about the student's past, present, and future.

The questions considered in the first phase are simple answers, but some have been derived into different branches to find more concrete answers. Thus, not all of the questions have been taken into account for the new weighting, resulting in 34 weighted questions. Some, such as "the place of study", were discarded because they were considered difficult to interpret and no bibliographical references were found to support their inclusion. Likewise, the use of self-regulation strategies has also been excluded from weighting due to the difficulty of collecting data in the questionnaire and the existence of some papers that point out their lack of correlation. However, both the initial and derived questions have been left in the questionnaire because they are important as knowledge about the students for good tutorial intervention.

Finally, through the analysis of expert tutors, possible student responses are categorized, as each response will receive a different weight depending on the category it falls under. This categorization is carried out using a qualitative method with expert tutors from different centres and areas. To do this, a focus group was conducted where the different questions with their possible responses were shown, and the impact of each response on the potential risk of student dropout was discussed. These items are validated as the process is replicated and different iterations are performed.

Once students with a higher or lower risk of dropout are identified, tutors are responsible for analysing the situation and speaking with the student. In this way, they can determine if the potential risk of dropout is real and how they can assist the student. While it is true that tutors cannot fully control the student's situation, in cases where the student is experiencing frustration due to poor results or lack of good study techniques, the tutor can provide tools for improvement. They can also assist in organizing time, tasks, review classes, etc. Additionally, a pilot test for coaching with these skills is being developed to aid students in facing the different challenges that arise during their first year of university studies.

b) Questionnaire

The questionnaire created for the students requires answers to 53 questions in total. Of those questions, only 33 are weighted to obtain the critical value for the likelihood of dropout, and the remaining 19 provide answers with extra information for the tutors. In Table 1, the different questions in the questionnaire are presented, with those that are weighted marked in grey and numbered and those that are not weighted providing extra information and unnumbered. All these questions are divided into the 6 large blocks that we will work with.

Table 1 shows each of the questions asked in the questionnaire with their corresponding weighting and the block to which they belong. Tutors give a score of 1–10 to each of the 6 blocks, depending on how important they consider them. Subsequently, they rank the questions in each block from most to least relevant. Before the start of the 2022–2023 academic year, each variable was weighted by a total of 53 first-year tutors from different academic fields in three Spanish universities: La Salle-RLU, Deusto (Bilbao campus) and Comillas (Madrid campus).

For example, in the first dimension, we found two variables, so each tutor ranked them based on their perception of importance, giving a score of 2 to the most important variable and 1 to the least important one. If a dimension has 6 variables, the ranking will range from 6 (most important) to 1 (least important). For this reason, the average that arises from the tutors' ratings for each variable ranges from 1 to the highest number of variables in that dimension.

To calculate the probability of dropout on a scale of 0–10 for each student, first, the weighting of each of the 6 blocks is calculated. First, the mean score given by all tutors is calculated, and then the following calculation (equation (1)) is performed to obtain the block value out of a total of 10:

$$\text{Block weighting} = \frac{\textit{Average block rating x } 10}{\sum \textit{Average rating of the } 6 \textit{ blocks}} \quad (1)$$

Table 2
Example of calculating the probability from the questionnaire answers.

Questions	User 23	User 47	User 90	User 147
Personal Data				
Year of birth	2004	2002	2003	2002
Gender	Woman	Woman	Man	Man
University entrance				
How did you enter university?	Selectividad	Selectividad	Selectividad	Certificate of Higher Education
From which baccalaureate/Voc. Training/other degree?	Technological	Artistic	Technological	
You consider yourself to be a student of:	5	4	3	4
How do you usually study (books, notes, classmates, etc.)?	My teachers have always given me study materials	I create my own study material	I create my own study material	I create my own study material
Average mark of university access	12.768	8.8	7.55	7.96
In which country did you study the baccalaureate or the compulsory course?	Spain	Spain	Spain	Spain
Current data				
What grade are you attending?	Engineering	Digital Arts and Animation	Architecture and Building	Engineering
Where do you live during the course?	In the family home	In the family home	In a residence or college	In the family home
How long does it take you to get to university?	Between 30 and 45 min	Between 1 h and 2 h	Between 15 and 30 min	Between 30 and 45 min
Do you have a scholarship?	Yes	Yes	No	No
How would you describe your level of knowledge of basic computer tools?	3	5	4	5
Do you have a computer at home?	Yes, laptop	Yes, desktop	Yes, laptop	Yes, laptop and desktop
Do you have siblings or close friends who are or have been students on your course?	No	Yes	Yes	No
What is your relationship with classmates like?	4	5	5	3
Do you have anyone among your classmates whom you could ask for notes when you cannot go to the classroom?	Yes	Yes	Yes	Yes
Degree choice				
Do you feel integrated in the classroom group?	Yes	Yes	Yes	Yes
How confident/assured are you about your chosen degree?	5	5	5	5
Were the studies you are pursuing listed as your first choice?	Yes	Yes	No	Yes
The syllabus of the 1st that I am studying seems to me to …	It is excessively abstract.	It is excessively abstract.	It meets my expectations.	It meets my expectations.
Study habits				
Do you complete the tasks given to you?	5	5	5	3
Do you study and revise every day?	5	5	5	1
Where do you usually study?	At home	At home	At home	At home
Do you expect to join an academy?	No	No	No	Yes
How many days before an exam do you study?	More than two weeks before	Between 3 and 5 days	Between 1 and 2 days	Between 1 and 2 days
Time spent studying				
Do you have a job, paid or unpaid?	No	No	No	No
Do you think it will prevent you from following the course normally?	No	Sí	Sí	No
Do you carry out any activity to which you dedicate a significant amount of time, and which may take time away from your university studies?	No	Sí	No	No
How many hours do you plan to spend during the week studying?	35	16	23	5
Are there any other circumstances (family, illness, etc.) that might make it difficult for you to follow your studies?	No	No	No	No
How many hours do you plan to dedicate to study during the week?	Yes, I want to	Yes, I want to	Yes, I want to	Yes, I want to
Are there any classes that you already know you will not be able to attend?	I like to do my best	I like to do my best	I like to do my best	I like to do my best
Are you motivated to study sufficiently for the courses you have enrolled in?	No, I think it is more demanding than I imagined	Yes, I think that's what I imagined	Yes, I think that's what I imagined	Yes, I think that's what I imagined
Total	8.02117277	7.20254052	6.94068171	5.97523526

Following the previous example, for the first block, personal data, an average block rating of 5.67 has been obtained, which we multiplied by 10 as it is the global value of the defined scale. In the denominator, we have the sum of the averages of the 6 blocks, including the one we are calculating and the remaining 5 blocks (5.68 + 6.15+7.06 + 7.78+8.26 + 8.09), as shown in Table 1. The resulting value is 1.318.

Once the block value is obtained, the value of each of the 34 questions is calculated (equation (2)). To perform this calculation, the tutors' scores are first averaged. This value is obtained from the average of the rankings. Once the average score for the question is obtained, the following calculation is performed:

$$Question\ weighting = \frac{Block\ weighting\ x\ Average\ rating\ of\ the\ question}{\sum Average\ questions\ rating} \tag{2}$$

For the first question, age, an average of 1.83 was obtained. This value is obtained from the tutors' ranking of age and gender. A score of 2 is given to the variable considered most important and 1 to the least important one. Then, the average of each variable is calculated and multiplied by the block weight; in this case, since it is a value from the personal dimension, that value is used (1.318*1.83). This value is divided by the sum of the averages of both values, 1.83 (age) + 1.17 (gender). These calculations provide us with the final weight of the age value, 0.805.

Subsequently, the group of experts classified the responses according to their greater or lesser impact on dropout and assigned the corresponding weighting.

$$\text{Weighting of answers} = \frac{Question\ weighting}{Total\ numer\ of\ answers}\ x\ Impact\ number \tag{3}$$

Finally, the weighting of each value is divided among the different responses that can be obtained (equation (3)). Following the previous example, we can see that the age value is divided into three possible responses: over 25 years old, between 20 and 25 years old, and between 18 and 20 years old. These 3 responses are classified, with students between 18 and 20 years old having the highest weighting (3), students between 20 and 25 years old having medium weighting (2), and students over 25 having the lowest weighting (1). To obtain the weighting of each response, first, the division is made between the weighting of the value and the total number of responses (1.318/3), and then it is multiplied by the number assigned to each response. In the case of students between 18 and 20 years old, since the assigned value is 3, they obtain a result of 0.805, which coincides with the total value. For students between 20 and 25 years old, a value of 0.537 is obtained, and for students over 25, a value of 0.268 is obtained.

Table 2 displays an example of 4 users with their different responses and the resulting calculation using global weighting.

The results of the weights were classified and weighted based on the tutor's area of knowledge and the centre to which he or she belonged. In this way, we obtained an overall weighting and a weighting for each of the areas and centres. On the other hand, the different questions of the survey shown above were grouped in such a way that 6 blocks were created with the same total score; in this way, the impact of each of the parts of the project can be analysed in a simpler way. In addition to providing more information to the tutor who receives the results, he or she will not only receive the final weighting but also a breakdown of the 6 blocks.

c) Calculation of the risk of dropout

During the start of the 2022–2023 academic year, a total of 1742 students were newly enrolled in the three participating universities in the fields of Engineering, Business, Arts, Architecture, Social and Human Sciences, Health Sciences, and Law. The requirement to be able to take the questionnaire is to be a newly enrolled student in the selected degree program and to belong to the first year of the degree program. These students responded to the questionnaire presented in Table 1.

As we have mentioned before, each of the tutors from different areas and centres completes a questionnaire where they assign different weights to different questions, which we later group into blocks. Table 4 through 8 show the weights that the tutors assigned out of 10, divided by centres, areas, and centre areas. These weights are applied to the responses provided by each of the incoming students. We perform the calculation using four different weights, which are shown below, so that we can observe which weight is closer to the reality of the student.

- Overall weighting: the overall weighting made by all teachers is considered, regardless of the area or teaching centre.
- Weighting by area: the weighting of all the tutors of the student's teaching area is considered, regardless of the centre.
- Weighting by area and university centre: the weighting of all the tutors in the teaching area of the centre where the student is located is considered.

Table 3
Weighting ranges.

Risk of abandonment	Weighting range
Very low risk	>8
Low risk	7 to 8
High risk	6 to 7
Very high risk	<6

Table 4
Average weights per block and centre.

Weightings	Personal information	Access to University	Current data	Choice of university degree	Study habits	Study time
La Salle-RLU	1.451	1.344	1.714	1.840	1.831	1.821
Deusto	1.695	1.408	1.432	1.527	2.005	1.933
Comillas	0.982	1.541	1.655	1.906	1.986	1.929
Overall average	1.318	1.430	1.641	1.858	1.921	1.882
Inter-university deviation	0.362	0.101	0.149	0.202	0.096	0.064

Table 5
Average weights per block and teaching area.

Weightings	Personal information	Access to University	Current data	Choice of a university degree	Study habits	Study time
Architecture	1.610	1.314	1.653	1.801	1.780	1.843
Arts	1.648	1.429	1.538	1.978	1.538	1.868
Engineering	1.140	1.445	1.573	1.717	2.103	2.022
Business Administration and Management	1.280	1.377	1.715	1.981	1.884	1.763
Law	1.713	1.474	1.474	1.554	1.952	1.833
Social and Human Sciences	1.111	1.534	1.772	1.825	1.905	1.852
Health Sciences	0.645	1.613	1.720	2.151	1.935	1.935
Overall average	1.318	1.430	1.641	1.858	1.921	1.882
The deviation between knowledge areas	0.3826	0.0988	0.1095	0.1965	0.1754	0.0829

Table 6
Average weights per engineering block by educational institution.

Weightings	Personal information	Access to University	Current data	Choice of a university degree	Study habits	Study time
Engineering La Salle-RLU	1.024	1.496	1.850	1.732	2.087	1.811
Engineering Deusto	1.442	1.442	1.346	1.635	2.115	2.019
Engineering Comillas	0.932	1.366	1.429	1.801	2.112	2.360
Engineering average	1.140	1.445	1.573	1.717	2.103	2.022
Engineering deviation	0.272	0.065	0.270	0.084	0.016	0.277

Table 7
Average weights per block of business administration and management by the educational establishment.

Weightings	Personal information	Access to University	Current data	Choice of a university degree	Study habits	Study time
Business Administration and Management La Salle-RLU	1.524	1.190	1.762	2.000	1.762	1.762
Business Administration and Management Comillas	1.029	1.569	1.667	1.961	2.010	1.765
Business Administration and Management average	1.280	1.377	1.715	1.981	1.884	1.763
Business Administration and Management deviation	0.350	0.267	0.067	0.028	0.175	0.002

Table 8
Average weights per entitlement block per educational establishment.

Weightings	Personal information	Access to University	Current data	Choice of a university degree	Study habits	Study time
Law Deusto	1.943	1.374	1.517	1.422	1.896	1.848
Law Comillas	0.500	2.000	1.250	2.250	2.250	1.750
Law average	1.713	1.474	1.474	1.554	1.952	1.833
Law deviation	1.020	0.442	0.189	0.586	0.251	0.070

- Weighting by university centre: the weighting of all the tutors of the university centre where the student is, regardless of the area, is considered.

The weighting from 0 to 10 is used as an approximation of the scoring system used in the degrees where the predictive system is applied. Unlike primary or secondary education, where the Spanish system has changed from 1 to 4, in the case of working in that area, the scale would be worked on accordingly. This scoring scale will be subjected to validation during the following years of replication of the system, using the obtained results to validate the designed scale.

d) Classification of the risk of dropout

The result obtained from the weighting of each of the responses provided by the student is divided into four bands. The bands determining the critical risk of dropout are shown in Table 3, which shows the different ranges.

It is divided into 4 ranges based on the results, as shown in Fig. 1. A Gaussian bell curve is created with the results, creating the standard deviation and normal distribution. This allows us to detect those students with a very high risk of dropout on the left side of the graph, with a score below 6, and those with very little risk of dropout on the right side of the graph, with a score above 8. The central part of the graph is divided into proportional parts to provide more information to the tutor. Thus, if they obtain a score between 6 and 7, we will say that they have a high probability of dropout, and between 7 and 8 a low probability.

4. Results

a) Weighting per block

Table 4 shows the different weightings according to centre. Initially, a significant difference between centres is already observed when we take into account personal data and degree choice.

If we analyse each of the areas of knowledge identified in the centres and grouped by common studies (see Table 5), we can observe significant differences from the data obtained from the weighting by the tutors, which indicates clearly differentiated student profiles. In this sense, a starting hypothesis in applying the questionnaire to students will be that the result will vary significantly between applying a global indicator or a specific one by area, the latter being much more precise.

These weights are also analysed according to the degree programme. In this case, in Tables 6–8, we can observe the different weights of the same area in different centres. Previously, the data with the highest amount of disagreement were personal data and degree choice, and when we focus on one of the areas where these values are more similar to each other, we can observe the tutor's weight has a similar value.

We can observe in Table 6 that there are discrepancies between the different centres and the weights. La Salle-RLU and Deusto consider the most relevant block for dropout to be study habits, totalling 2.087 and 2.115 out of 10, respectively. Meanwhile, at Comillas, the amount of time dedicated to studying is considered the most relevant block, with 2.360 out of 10.

In Table 7, we focus on the business area, where the most important blocks vary from those used with engineering students. There are discrepancies among different centres due to the perceptions of teachers. La Salle-RLU tutors consider the motivation of students with a chosen degree as the most important factor with 2 out of 10. In Comillas, the most important block for tutors is study habits, with 2.010 out of 10.

The area of law is where the greatest differences are found among different centres. Deusto tutors rated personal data as the most important factor in the final weighting, with 1.943 out of 10. In contrast, in Comillas, the motivation for the chosen degree and study habits are considered the most important with 2.250 each.

If we specifically analyse the data by teaching areas, we can see in Table 5 that there are discrepancies, but all tutors position personal information (1.318 out of 10), access to university (1.430 out of 10), and current data (1.641 out of 10) as the least important

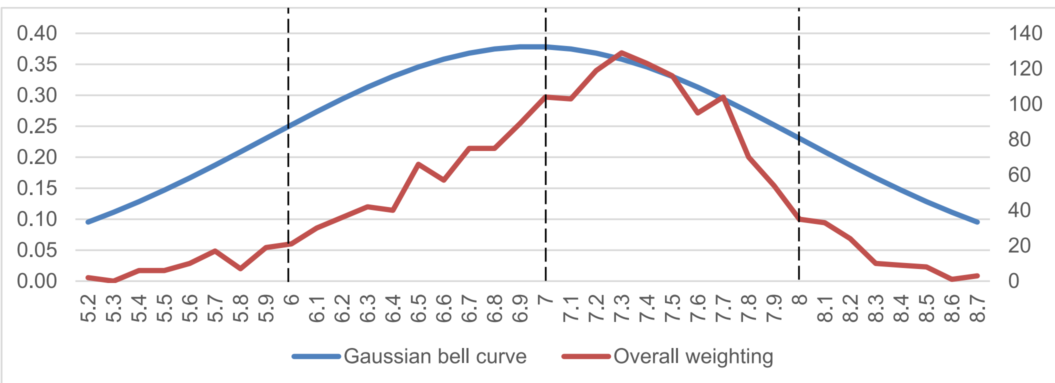

Fig. 1. Gaussian bell curve and global probability of dropout for range creation.

values for detecting student dropout. In the architecture and arts majors, the blocks highlighted by their tutors are the choice of degree with 1.801 and 1.978, respectively, the vocation of those students for that major, and the time dedicated to the major with 1.843 and 1.868, respectively. If we focus on engineering students, one of the most important blocks is also study time, with 2.022 out of 10, coinciding with the vast majority of the areas studied. In this case, there are differences with the two majors mentioned earlier, as their tutors consider study habits to be more important, weighting the block with 2.103. In degrees such as Business Management and Administration, the tutors indicate that the study habit blocks are relatively important with 1.884, along with the student's vocation for that major with 1.981. In the case of Social Science degrees, such as Law, Philosophy, Education, Social Work, inter alia, which are more theoretical degrees, the most relevant blocks are those related to study habits with approximately 1.9 and time dedicated to them with approximately 1.8. Finally, in the majors related to health sciences, it is observed that the block where the tutors give the most weight is to vocation, obtaining the highest weighting of the entire table with 2.151, followed by study time and study habits, which obtain an equal weighting of 1.935. In this case, the weighting of personal data is significantly reduced since the general average is 1.318 and drops to 0.645.

As seen, the most important blocks for the tutors of the specific areas studied are the blocks that receive the most weight in the general weighting. However, there are differences among the tutors of the different areas, which demonstrates the importance of different weights depending on the student's profile.

b) Comparison of the different weights

Once the 4 ranges have been obtained, students are classified from highest to lowest risk of dropout. In Fig. 2, we can observe the classification of the risk of dropout according to the degree to which they are enrolled.

To analyse the potential risk of dropout among students and compare the different percentages obtained based on the weighting used, Table 9 is created. This table shows the percentage of dropouts classified from very high to very low probability, vertically according to the area where those students are located, and horizontally according to the different weights used, which take into account overall weighting, weighting by area, weighting by area and university centre, and weighting by university centre.

As seen in Table 9, there are differences between the generic and specific weights. In the areas of Engineering, Architecture, Social and Human Sciences, Health Sciences, and Business Management and Administration, it is already observed that if the criteria of specific tutors are taken into account, the risk of dropout is higher than if the generic weighting is applied. This increases the number of students with a high risk of dropout.

In the specific case of engineering, we can identify that 39.224% of the total students in that area are classified with a low risk of dropout under the general weighting; however, if we observe the specific weighting of the area, that percentage drops to 14.943%. In contrast, the high risk of dropout sees an increase in the number of students, initially at 55.172% and under the specific weighting system, 69.109%. In the very high risk of dropout, we can observe the same occurrence, with the percentage increasing from 4.454% to 15.230%.

In degrees such as Law, a Gaussian bell curve is created, where the greatest weight is in the central zone, that of high risk with 60.655%. In the case of this area, there are few differences between the general and specific weights, and we can see that the low risk of dropout ranges from 36.066% to 32.787% and the high risk of dropout ranges from 60.655% to 64.344%, with these values being very similar to the general values. Differences can be appreciated when analysing the area-related column in the specific centre, where the low risk of dropout goes from 36.066% under the general weighting system to 13.115%. On the other hand, the high risk of dropout and the very high risk of dropout are also modified from 60.655% to 68.852% and from 3.279% to 17.623%, respectively.

Finally, the other area with different centre-specific behaviour is Arts, where under general weighting, the greatest likelihood of dropout is in the high risk category at 61.403%. If the weighting of the area tutors is taken into account, it can be observed that the high category with 40.351% of students is translated into either low risk with 42.105% or very high risk of dropout with 15.789%. In this

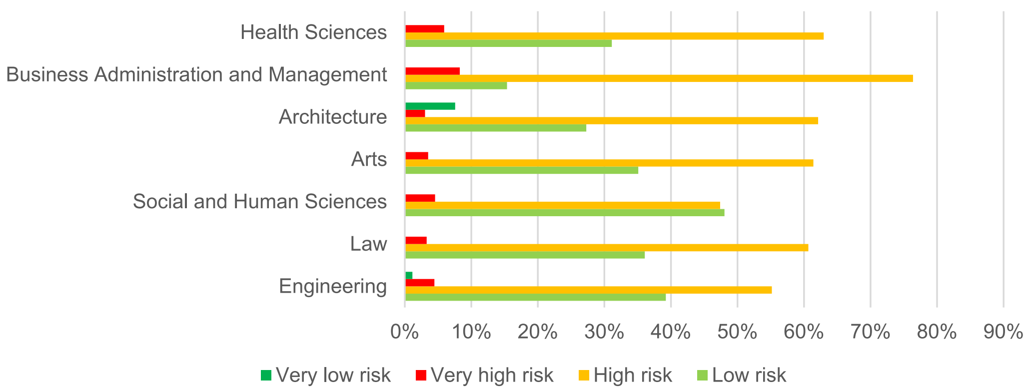

Fig. 2. Overall weight of each knowledge field.

Table 9
Results of the weights between the different degrees and university centres.

Risk of abandonment	Overall weighting	Weighting by area	Weighting by area and university centre	Weighting by a university centre	Deviation
Engineering					
Very low risk	1.149%	0.718%	0.862%	1.006%	0.185%
Low risk	39.224%	14.943%	14.368%	38.362%	13.942%
High risk	55.172%	69.109%	63.793%	56.034%	6.638%
Very high risk	4.454%	15.230%	20.977%	4.598%	8.183%
Law					
Very low risk	0.000%	0.000%	0.410%	0.000%	0.205%
Low risk	36.066%	32.787%	13.115%	36.066%	11.038%
High risk	60.655%	64.344%	68.852%	59.836%	4.117%
Very high risk	3.279%	2.869%	17.623%	4.098%	7.122%
Social and Human Sciences					
Very low risk	0.000%	0.000%		0.000%	0.000%
Low risk	48.052%	15.584%		49.351%	19.131%
Medium risk	32.468%	33.117%		31.169%	0.992%
High risk	47,403%	72,078%		46,104%	14,636%
Very high risk	4.545%	12.338%		4.545%	4.499%
Arts					
Very low risk	0.000%	1.754%		0.000%	1.013%
Low risk	35.088%	42.105%		38.596%	3.509%
High risk	61.403%	40.351%		57.895%	11.279%
Very high risk	3.509%	15.789%		3.509%	7.090%
Architecture					
Very low risk	7.576%	3.030%		7.576%	2.624%
Low risk	27.273%	19.697%		30.303%	5.463%
High risk	62.121%	68.182%		59.091%	4.629%
Very high risk	3.030%	9.091%		3.030%	3.499%
Business Administration and Management					
Very low risk	0.000%	0.000%	0.000%	0.000%	0.000%
Low risk	15.354%	16.142%	10.630%	18.504%	3.302%
High risk	76,378%	75,984%	73,228%	73,228%	1712%
Very high risk	8.268%	7.874%	16.142%	8.268%	4.007%
Health Sciences					
Very low risk	0.000%	0.000%		0.000%	0.000%
Low risk	31.111%	11.481%		32.963%	11.904%
High risk	62,963%	75,926%		62,963%	7484%
Very high risk	5.926%	12.593%		4.074%	4.480%

way, students of the Arts can be profiled more effectively, since by tending towards a higher or lower band, they provide more information about their risk and probability of dropping out of the degree.

5. Discussion

The tables shown previously (Tables 4–8) demonstrate the construction of an instrument that measures the risk of student dropout based on the weighting of personal variables by tutors. The variables that receive the greatest weight are those related to the motivation of the student in the selected degree (1.858), study habits (1.921), and the time they can dedicate to their studies (1.882), as we can observe in Table 4. These three dimensions add up to more than half of the final score among the three factors, thus answering RQ1.

These three dimensions consist of several values, which also carry varying degrees of importance according to the tutors. Degree choice (1.858) contains three values with similar weightings, with the security of the chosen degree (0.629) having the highest weight and therefore the most importance, followed by the other two variables with similar weights: whether it was the first choice in degree selection (0.585) and perception of the curriculum (0.596).

Table 10
Z test for the mean of the sample.

	Comparison 1		Comparison 2		Comparison 3	
	Overall	Area	Overall	University centre	Overall	Area and university centre
Media	7.137154	7.11256	7.137154	7.121547	7.11311307	7.1021846
Variance (known)	0.365832	0.373473	0.365832	0.378356	0.3600329	0.44273274
Observations	1738	1738	1738	1738	1192	1192
z	1.192465		0.754254		0.42111726	
P (Z ≤ z) one-tail	0.116539		0.225348		0.33683473	
Critical value of z (one-tail)	1.644854		1.644854		1.64485363	
Critical value of z (two-tail)	0.233079		0.450697		0.67366946	
Critical value of z (two-tail)	1.959964		1.959964		1.95996398	

If we focus on study habits (1.921), we can observe differences in the weighting of the internal variables. Tutors consider the most important variable in this dimension to be whether students study and review their subjects daily (0.517), which accounts for over a quarter of the final weight, followed by whether they complete the tasks assigned to them (0.51). The sum of these two variables accounts for more than half of the final score of the dimension, making them the most relevant and important variables for monitoring students during their degree.

Finally, with the highest score obtained in the last dimension, study time (1.882) reveals that the variables have relatively similar values. The variable with the highest importance is motivation to study (0.333), followed by the existence of any family situation that may affect their studies (0.281) and the hours they plan to dedicate to their studies (0.276).

Next, the comparison of the results using different weightings, using a z test, provides data shown in Table 10. In the process of applying the questionnaire to student responses, it becomes clear that different approaches can be applied. That is, the average weights of all tutors, regardless of the field of knowledge or centre, can be applied. Alternatively, the average weights from the tutors in the same centre can be applied, or even the average weights per area of study/degree in which the potential dropout risk is being studied. To analyse whether the differences obtained from the different averages of weights are significant, we applied Student's z test, which is suitable for the conditions of the samples we have. On the one hand, the variances of the samples are known, the samples are independent of each other, and the sample size is greater than 30 with a normal distribution.

Student's z test was used, with the null hypothesis (H0) that there are differences between the application of the different weighting systems. The statistical significance p is 0.116539 for *Comparison 1* (Overall weighting and Weighting by area), 0.225348 for *Comparison 2* (Overall weighting and Weighting by a university centre) and 0.3368 (Overall weighting and Weighting by area and university centre), which all exceed the threshold of 0.05, meaning there is a very low probability that the different weighting systems produce different results (Table 10). The null hypothesis is rejected, such that there are no significant differences in the weighting of the main variables by the tutors using area of knowledge, university and/or region to affect the calculation of the dropout probability, thus answering RQ2.

However, in the process of categorizing the results obtained through different weighting methods, the specific tutors for each area observed how those small differences from the general survey that are observed in the weighting by area of discipline seem to be more likely to approximate the reality of their students, even though the differences, as it has been suggest, are not statistically significant. This can be observed in Table 11.

This subjective perception of the tutors, according to the weighting applied by knowledge domain, seems to fit better than the rest of the weighting systems, providing us with a path to explore in future iterations. Notably, qualitative approaches (tutors' perception) appear to challenge or compromise the quantitative data obtained. In this sense, mixed approaches for user-focused research have already demonstrated their validity and effectiveness, opening a line of research that compares this perception with semiautomatic prediction systems [43–45].

We can observe that the percentages in the overall weighting system and in the weighting of the specific centre are the same, while the application of the weighting system by area and centre is less similar than the rest. In the case of applying weighting systems by area, a difference in accuracy of 1% can be observed. These values will be corroborated at the end of the course as the final results of the students will be obtained.

The results obtained from the designed survey instrument show that a predictive process can be tailored to personal variables of the student at the beginning of university that configure his or her potential risk of dropping out. This risk can be treated both objectively, based on the value obtained, or subjectively by the tutor, who has experience with the behaviour of students based on their academic performance, especially in the first courses.

The study has shown that by collecting these variables, the risk predicted by the survey instrument is close to the perception of the tutors, so it can be a very useful tool for teachers or tutors with less teaching experience to identify students at risk. Moreover, as seen from the data, the tool is not only predictive and functional in the early stages of the course but also adapts without significant differences to any type of degree programme, whether it is in the technical field or in the academic field.

6. Conclusions

The research presented here demonstrates that it is possible to define an entry profile for different groups of students and identify the variables needed to predict dropout. Thanks to the survey instrument created, tutors were able to reduce their initial effort and help students anticipate their work. Initially, tutors did not receive information about their mentees until the first tutorial meeting. The questionnaires help the tutors receive data at the beginning of the year to speed up their work and detect at-risk students as soon as possible. These aspects are essential in a university environment because it has been proven that the proportion of students who drop out increases considerably for those at the beginning of their studies, even in the first few months.

The combination of preexisting indicators, in some cases adopted from other studies, and improved through a qualitative Delphi

Table 11
Comparison between teachers' perception in the first semester and predictions.

	Overall weighting	Weighting by area	Weighting by area and university centre	Weighting by a university centre
Same	81%	82%	76%	81%
Medium	14%	13%	18%	14%
Opposite	5%	5%	6%	5%

process with tutors, provides an instrument that in initial assessments by tutors identifies potential students at risk of dropout in any field of knowledge with high accuracy (above 80%). This fact may help to replicate the study in other fields and/or further refine whether the instrument's behaviour is consistent for public, private, and/or any branch of knowledge and/or geographic scope.

The study was carried out in three different geographical areas, with three private universities and more than 10 different degrees, covering scientific-technical, humanistic, and social areas. In this sense, the results of the study would be applicable to any state university with an existing academic tutoring process. Undoubtedly, one of the potential weaknesses of the study is its limitation to these private institutions. The medium-term goal is to scale the proposal to public centres throughout Spain and Europe through collaborations in Erasmus + projects.

The process, as we have mentioned, is based on an iterative methodology of data exploration, and the reliability of the instrument will be confirmed over the next two years through internal measures and comparisons of its application. Over the next few academic years, the weights of the indicators will be adjusted, leading to a better approximation to reality, as well as the search for a semi-automatic format that allows the survey instrument to generate predictions of characteristic behaviour. In this sense, the main contribution of the article compared to previous studies is the creation of a tool that provides an indicator value that predicts the potential risk of student dropout, as a function of relating all the studied variables.

Evaluating the results obtained from different centres, the most influential data from students who drop out prematurely are the choice of degree, study habits, and the time dedicated to study. Continuing with the analysis carried out in recent years, the data collected are expanded and individualized weighting systems are made instead of generic ones, and by repeatedly evaluating this process over the years, this project establishes that it is possible to control dropouts that may occur due to frustration, lack of motivation, or lack of knowledge of the chosen degree.

The main limitations of the proposal are identified as future challenges to be addressed. First, it is necessary to delve into the comparison between the data obtained from the survey and the tutors' perception. As discussed, while there are initially no significant differences between the use of global, centre-based, or discipline-based weights, subjective differences by field of study have been observed. Certain tutors in very specific fields such as architecture or engineering identify field averages as more in line with their observed reality and knowledge of their students. This aspect is vital because it will define the tools and actions that tutors can apply to mitigate the risk of dropout. In this sense, pilot experiences are already being carried out in the universities under study to apply coaching-derived processes to students identified by the initial questionnaire. The success of the applied actions will need to be verified in subsequent academic years.

Both the success of the data analysis and tutor activities that will validate the instrument will be based on the number of dropouts at the end of the year, understood as a success if that number has been initially identified by the system, assuming that some students initially identified as high risk will not drop out thanks to the interventions of the tutors and their work throughout the year. Likewise, with this end-of-year monitoring process and through temporal repetitions, the following years will be able to refine the weights and ultimately the instrument for each of the disciplines and centres. The iterative nature of the proposal is another limitation of the study that the project itself will solve throughout the 2023–2024 academic year, based on its repetitive and iterative methodology. With potential adjustments to the questionnaire weights based on the validation of the data, a more reliable tool will be obtained for predicting the risk of dropout.

Similarly, as a challenge linked to the limitation of working with the perceptions of tutors and students, we are already investigating the use of artificial intelligence tools and semiautomatic predictive algorithms to compare the results obtained from the survey instrument with automatic analysis of tutor monitoring/assessment. This aspect is vital, as it will improve the tutor's workflow without the need for an initial predictive process but by directly using the data returned by the system to establish a strategy to monitor, intervene, and support students.

Funding

This research was funded by the project "Academic Analytics applied in the study of the relationship between the initial profile of undergraduate students and early drop-out rates in order to improve tutorial support processes [ASPA4DOR]", granted at the VIII Call of ACM (Aristos Campus Mundus) Research Projects—2022, with the grant number: [ACM2022_04].

Ethics statement

The research presented, as well as the design, collection, and management of its data, has been POSITIVE evaluated and APPROVED, by the Ethics Committee of the Ramon Llull University with the file number CER URL_2021_2022_014 (with date July 18, 2022), and supplementary by the Deusto University Ethics Committee of Research, with Reference: ETK-41/21–22 (with date July 27, 2022), and also with the positive validation of the Ethics Committee of the Pontific University of Comillas with reference 2022/48 of data August 30, 2022.

Author contribution statement

Alba Llauró: Conceived and designed the experiments; Performed the experiments; Analysed and interpreted the data; Wrote the paper.

David Fonseca: Conceived and designed the experiments; Analysed and interpreted the data; Wrote the paper.

Susana Romero: Conceived and designed the experiments; Wrote the paper.

Marian Aláez; Jorge Torres Lucas: Contributed reagents, materials, analysis tools or data; Wrote the paper.
María Martínez Felipe: Performed the experiments; Wrote the paper.

Data availability statement

Data included in article/supplementary material/referenced in article.

Additional information

Supplementary content related to this article has been published online at [URL].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J.-F. Espinosa-Castro, J. Hernandez-Lalinde, L.M.M. Castro, Estrategias de permanencia universitaria, Arch. Venez. Farmacol. Ter. 39 (2020) 88–97.
[2] R. Garcia-Ros, F. Perez-Gonzalez, F. Cavas-Martinez, J.M. Tomas, Social interaction learning strategies, motivation, first-year students' experiences and permanence in university studies, Educ. Psychol. 38 (2018) 451–469.
[3] T. Pino-Vera, E. Cavieres-Fernandez, J.-A. Munoz-Reyes, T. Pino-Vera, E. Cavieres-Fernandez, J.-A. Munoz-Reyes, Los factores personales e institucionales en el sentido de pertenencia de estudiantes chilenos a lo largo de sus estudios superiores, Rev. Iberoam. Educ. Super. 9 (2018) 24–41, https://doi.org/10.22201/iisue.20072872e.2019.25.337.
[4] S.F. Hamilton, Raising standards and reducing dropout rates, Teach. Coll. Rec. 87 (1986) 410–429.
[5] M. Esteban, A. Bernardo, L. Rodriguez-Muniz, Persistence in university studies: the importance of a good start, Aula Abierta 44 (2016) 1, https://doi.org/10.17811/rifie.44.2016.1-6.
[6] A. Constante-Amores, M. Fernandez-Mellizo, E.F. Martinez, E.N. Asencio, Factores asociados al abandono universitario, Educ. XX1 24 (2021) 17–44.
[7] O. Gutierrez-Aguilar, J. Bautista-Lopez, E. Sanchez-Meza, R. Falcon-Huamani, Y. Condori-Surco, Y. Tomaylla-Quispe, The virtual teaching process and academic self-efficacy in learning and the intention to drop out university studies in times of Covid-19, in: 2021 XVI Lat. Am. Conf. Learn. Technol. LACLO, IEEE, 2021, pp. 376–382.
[8] L.A. Katt-Morales, M. Dominguez-Campomanes, E. Gomez-Manuel, E. Morales-Reyes, Strategies to Promote University Permanence in Times of COVID-19, J. Univ. Policies, 2021, pp. 5–14.
[9] P. Oreopoulos, R. Ford, Keeping College options open: a field experiment to help all high school seniors through the College application process, J. Pol. Anal. Manag. 38 (2019) 426–454, https://doi.org/10.1002/pam.22115.
[10] D.A. Filva, M.A. Forment, F.J. Garcia-Penalvo, D.F. Escudero, M.J. Casan, Clickstream for learning analytics to assess students' behavior with Scratch, Future Generat. Comput. Syst. 93 (2019) 673–686, https://doi.org/10.1016/j.future.2018.10.057.
[11] F.J. Garcia-Penalvo, Avoiding the dark side of digital transformation in teaching. An institutional reference framework for eLearning in higher education, Sustainability 13 (2021) 2023, https://doi.org/10.3390/su13042023.
[12] A.H. Duin, J. Tham, The current state of analytics: implications for learning management system (LMS) use in writing pedagogy, comput, Composer 55 (2020), 102544, https://doi.org/10.1016/j.compcom.2020.102544.
[13] A.Y.Q. Huang, O.H.T. Lu, J.C.H. Huang, C.J. Yin, S.J.H. Yang, Predicting students' academic performance by using educational big data and learning analytics: evaluation of classification methods and learning logs, Interact. Learn. Environ. 28 (2020) 206–230, https://doi.org/10.1080/10494820.2019.1636086.
[14] R.S. Baker, T. Martin, L.M. Rossi, Educational data mining and learning analytics, in: Wiley Handb. Cogn. Assess., John Wiley & Sons, Ltd, 2016, pp. 379–396, https://doi.org/10.1002/9781118956588.ch16.
[15] H. Waheed, S.-U. Hassan, N.R. Aljohani, J. Hardman, S. Alelyani, R. Nawaz, Predicting academic performance of students from VLE big data using deep learning models, Comput. Hum. Behav. 104 (2020) 106189, https://doi.org/10.1016/j.chb.2019.106189.
[16] P. Baepler, C. Murdoch, Academic analytics and data mining in higher education, Int. J. Scholarsh. Teach. Learn. 4 (2010), https://doi.org/10.20429/ijsotl.2010.040217.
[17] S. Palmer, Modelling engineering student academic performance using academic analytics, Int. J. Eng. Educ. 29 (2013) 132–138.
[18] D. Perez-Jorge, M. del C. Rodriguez-Jimenez, E. Arino-Mateo, F. Barragan-Medero, The effect of covid-19 in university tutoring models, Sustainability 12 (2020) 8631.
[19] C. Chabbott, M. Sinclair, SDG 4 and the COVID-19 emergency: textbooks, tutoring, and teachers, Prospects 49 (2020) 51–57.
[20] C. Johns, M. Mills, Online mathematics tutoring during the COVID-19 pandemic: recommendations for best practices, Primus 31 (2021) 99–117.
[21] S.E. Volet, P.D. Renshaw, Cross-cultural differences in university students' goals and perceptions of study settings for achieving their own goals, High, Educ. Next 30 (1995) 407–433.
[22] S. Rodriguez Espinar, M. Alvarez Gonzalez, Manual de tutoría universitaria recursos para la acción, Editorial Octaedro : Universitat de Barcelona, Institut de Ciències de l'Educació, Barcelona, 2012.
[23] G. Villa Fernandez, J.A. Montero Morales, A. Llauro Moliner, Educational coaching applied to group tutoring sessions: an experience with first-year engineering students, in: Eighth Int. Conf. Technol. Ecosyst. Enhancing Multicult., Association for Computing Machinery, New York, NY, USA, 2020, pp. 339–344, https://doi.org/10.1145/3434780.3436588.
[24] J. Nielsen, The usability engineering life cycle, Computer (1992), https://doi.org/10.1109/2.121503.
[25] Design methods for developing services, (n.d.). https://www.designcouncil.org.uk/our-work/skills-learning/resources/design-methods-developing-services/ (accessed January 23, 2023).
[26] R.S. Adams, C.J. Atman, Cognitive processes in iterative design behavior, in: FIE99 Front. Educ. 29th Annu. Front. Educ. Conf. Des. Future Sci. Eng. Educ. Conf. Proc. IEEE Cat No99CH37011, 1999, https://doi.org/10.1109/FIE.1999.839114, 11A6/13-11A6/18 vol.1.
[27] B. Gros, E. Durall, Retos y oportunidades del diseño participativo en tecnología educativa, Edutec Rev. Electrónica Tecnol. Educ. (2020) 12–24, https://doi.org/10.21556/edutec.2020.74.1761.
[28] A. Llauro, D. Fonseca, D. Amo-Filva, S. Romero, M. Alaez, J. Torres Lucas, M. Martinez Felipe, Academic analytics applied in the study of the relationship between the initial profile of undergraduate students and early drop-out rates. Defining the variables of a predictor instrument, in: F.J. Garcia-Penalvo, A. Garcia-Holgado (Eds.), Proc. TEEM 2022 Tenth Int. Conf. Technol. Ecosyst. Enhancing Multicult., Springer Nature, Singapore, 2023, pp. 982–990, https://doi.org/10.1007/978-981-99-0942-1_103.
[29] R.C. Munoz, I.M. Greca, V.T. Gomez, M.C. Fernandez, A.M.L. Palma, Impacto de un Plan de Acción Tutorial universitario: resultados académicos, implicación y satisfacción, REDU Rev. Docencia Univ. 12 (2014) 323–342, https://doi.org/10.4995/redu.2014.5626.

[30] J.M. Braxton, J.F. Milem, A.S. Sullivan, The influence of active learning on the College student departure process, J. High. Educ. 71 (2000) 569–590, https://doi.org/10.1080/00221546.2000.11778853.

[31] V. Realinho, J. Machado, L. Baptista, M.V. Martins, Predicting student dropout and academic success, Data 7 (2022) 146, https://doi.org/10.3390/data7110146.

[32] J.C. Almenara, A.I. Moro, Empleo del método Delphi y su empleo en la investigación en comunicación y educación, Edutec Rev. Electrónica Tecnol. Educ. (2014) a272, https://doi.org/10.21556/edutec.2014.48.187. –a272.

[33] A. Llauro, D. Fonseca, E. Villegas, M. Alaez, S. Romero, Educational data mining application for improving the academic tutorial sessions, and the reduction of early dropout in undergraduate students, in: Ninth Int. Conf. Technol. Ecosyst. Enhancing Multicult. TEEM21, 2021, pp. 212–218.

[34] M. Esteban Garcia, A.B. Bernardo Gutierrez, L.J. Rodriguez-Muniz, Persistence in university studies: the importance of a good start, Aula Abierta 44 (2016) 1–6, https://doi.org/10.1016/j.aula.2015.04.001.

[35] I. Sandoval-Palis, D. Naranjo, J. Vidal, R. Gilar-Corbi, Early dropout prediction model: a case study of university leveling course students, Sustainability 12 (2020) 9314, https://doi.org/10.3390/su12229314.

[36] E.E.M. Meens, A.W.E.A. Bakx, T.A. Klimstra, J.J.A. Denissen, The association of identity and motivation with students' academic achievement in higher education, Learn, Individ. Differ. 64 (2018) 54–70, https://doi.org/10.1016/j.lindif.2018.04.006.

[37] V. Germeijs, K. Verschueren, High school students' career decision-making process: consequences for choice implementation in higher education, J. Vocat. Behav. 70 (2007) 223–241, https://doi.org/10.1016/j.jvb.2006.10.004.

[38] C. Vieira, I. Vieira, L. Raposo, Distance and academic performance in higher education, Spatial Econ. Anal. 13 (2018) 60–79, https://doi.org/10.1080/17421772.2017.1369146.

[39] I. Fomunung, M. Porter, C. Silver, Increasing the success rates of engineering students after transferring into four year colleges from community colleges: it's much more than dollars. https://peer.asee.org/increasing-the-success-rates-of-engineering-students-after-transferring-into-four-year-colleges-from-community-colleges-it-s-much-more-than-dollars, 2022 (accessed March 28, 2023).

[40] B.M. Kehm, M.R. Larsen, H.B. Sommersel, Student dropout from universities in Europe: a review of empirical literature, Hung. Educ. Res. J. 9 (2019) 147–164, https://doi.org/10.1556/063.9.2019.1.18.

[41] A. Abu Saa, M. Al-Emran, K. Shaalan, Factors affecting students' performance in higher education: a systematic review of predictive data mining techniques, Technol. Knowl. Learn. 24 (2019) 567–598, https://doi.org/10.1007/s10758-019-09408-7.

[42] P.J.A.C. van der Zanden, E. Denessen, A.H.N. Cillessen, P.C. Meijer, Domains and predictors of first-year student success: a systematic review, Educ. Res. Rev. 23 (2018) 57–77, https://doi.org/10.1016/j.edurev.2018.01.001.

[43] D. Fonseca, E. Redondo, S. Villagrasa, Mixed-methods research: a new approach to evaluating the motivation and satisfaction of university students using advanced visual technologies, Univers. Access Inf. Soc. 14 (2015) 311–332, https://doi.org/10.1007/s10209-014-0361-4.

[44] D. Fonseca, J. Cavalcanti, E. Pena, V. Valls, M. Sanchez-Sepulveda, F. Moreira, I. Navarro, E. Redondo, Mixed assessment of virtual serious games applied in architectural and urban design education, Sensors 21 (2021) 3102, https://doi.org/10.3390/s21093102.

[45] E. Villegas, D. Fonseca, E. Pena, P. Bonet, S. Fernandez-Guinea, Qualitative assessment of effective gamification design processes using motivators to identify game mechanics, Sensors 21 (2021) 2556, https://doi.org/10.3390/s21072556.