Journal of Management and Business Education

JMBE

www.journaljmbe.com ISSN: 2605-1044 Published by Academia Europea de Dirección y Economía de la Empresa.

Grades for glory? The impact of grade appeasement on business school teacher evaluations

¿Calificaciones para la gloria? El impacto de la complacencia de notas en las evaluaciones de los docentes de escuelas de negocios

Jorge Villagrasa*
https://orcid.org/0000-0001-9166-5937 (ORCID iD)
EDEM Business School (Spain)

Colin Donaldson
https://orcid.org/0000-0001-6224-4295 (ORCID iD)
EDEM Business School (Spain)

Ángel Soler
https://orcid.org/0000-0003-3721-6692 (ORCID iD)
University of Valencia (Spain)

Villagrasa, J. ; Donaldson, C. ; & Soler, A. (2024). Grades for glory? The impact of grade appeasement on business school teacher evaluations. *Journal of Management and Business Education*, 7(2), 303-323. https://doi.org/10.35564/jmbe.2024.0017
*Corresponding author: jvillagrasa@edem.es
Language: English
Received: 9 August 2023 / Accepted: 26 May 2024

Funding. The authors received no financial support for the research, authorship, and/or publication of this article.

Ethical Statement. The authors confirmed that data collection for the research was conducted anonymously and analyzed in an aggregate format, with no possibility of identifying the participants of this study.

Declaration of conflicting interests. The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CRediT author statement. Jorge Villagrasa: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Supervision, Validation, Visualization, Writing – original draft preparation, Writing – reviewing & editing. Colin Donaldson: Data curation, Formal Analysis, Investigation, Methodology, Supervision, Visualization, Writing – reviewing & editing. Angel Soler: Funding acquisition, Project administration, Resources, Validation, Writing – reviewing & editing

ABSTRACT

Educational institutions have had the pending task of knowing what really happens in their classrooms, which professors make a real difference, which students would need help to improve their performance, and even more importantly, how to manage the enormous amount of information gathered each year. To address this problem, in this paper we propose the use of retrospective data from 210 undergraduate courses in Spain, which equals 1,320 European credits, 10,500

students, and four complete academic years (18/19, 19/20, 20/21, 21/22) to generate a monitoring tool or automatic dashboard with which to try to optimize the operation of such institutions and help them to make (even more) optimal decisions. To do this, all courses were grouped into ten typologies or categories: (1) if they were 'soft skills'/'hard skills' courses, (2) if they were taught in the 1st, 2nd, 3rd or 4th academic year, (3) if the professor was male or female, (4) whether the course was taught in English or not, (5) whether the professor was internal and permanent staff of the university or an associate professor with a temporary contractual relationship with the institution, (6) whether his/her main professional activity was teaching, therefore being more academic-oriented or this was rather a hobby and a complementary professional activity, therefore being more practical-oriented, (7) their age, (8) if several professors shared the course or only one taught it, (9) if it was a semester or annual course, (10) and if it was taught during the Covid-19 period or outside of it. Our results show that, despite the existence of numerous types of methodologies to achieve the desired student motivation, there would only be one category that would generate significant differences in the grades of the students: if they were 'soft skills'/'hard skills' courses. However, on the contrary, there would be much more types of courses (and professors) which would affect the preferences and valuation of the student to the teaching received (and effectively paid by him/her). That is, despite reaching the same academic results, students do value one type of professors better than others. Likewise, through this work it is intended to demonstrate the non-existence of the "false belief" that better evaluations received by professors are since they give higher grades to their students, regardless of its dispersion. In this way, it is intended to prevent possible atypical behaviors by professors. Thus, this paper gives light to (re)consider which variables should actually matter in educational institutions, analyzing its teaching-learning processes, management and even the retention of its key players.

Keywords. educational methodologies, student motivation, student preferences, professor evaluations, optimization

RESUMEN

Las instituciones educativas han mantenido pendiente la tarea de saber qué sucede realmente en sus aulas, qué profesores/as marcan una diferencia real, qué alumnos/as necesitarían ayuda para mejorar su desempeño y, más importante aún si cabe, cómo poder manejar la enorme cantidad de información generada año a año. Para abordar este problema, en este artículo se propone el uso de datos retrospectivos de 210 asignaturas de grado en España, que equivalen a 1.320 créditos europeos, 10.500 estudiantes y cuatro cursos académicos completos (18/19, 19/20, 20/21, 21/22) a la hora de generar una herramienta de monitorización o cuadro de mandos automático con el que intentar optimizar el funcionamiento de dichas entidades y ayudarlas a tomar decisiones (aún más) óptimas. Para ello se agruparon todas las asignaturas en diez tipologías o categorías: (1) si eran asignaturas 'soft skills'/'hard skills', (2) si se impartían en 1º, 2º, 3º o 4º curso académico, (3) si el profesor era hombre o mujer, (4) si la asignatura se impartía en inglés o no, (5) si el profesor era personal interno y permanente de la universidad o profesor asociado de la misma, manteniendo un contrato temporal, (6) si su principal actividad profesional era la de dar clase, y por tanto tenía un perfil más académico o si por el contrario el dar clase era más bien un hobby y una actividad profesional complementaria, teniendo por tanto un perfil más orientado a la práctica y a la industria, (7) su edad, (8) si varios profesores compartían la asignatura o era uno solo el que la impartía, (9) si la asignatura era semestral o anual, (10) y si se impartía durante el periodo del Covid-19 o fuera de él. Nuestros resultados muestran que, a pesar de la existencia de numerosos tipos de metodologías para lograr la deseada motivación en los estudiantes, tan solo una categoría generaría diferencias significativas en las notas de los estudiantes: si son asignaturas 'soft skills'/'hard skills'. Sin embargo, por el contrario, existirían muchos más tipos de asignaturas (y profesores/as) que afectarían a las preferencias y valoración del alumno/a acerca

de la enseñanza recibida (y efectivamente pagada por él/ella). Es decir, a pesar de alcanzar unos mismos resultados académicos, los alumnos/as sí valoran mejor a un tipo de profesores/as que a otros. Asimismo, a través de este trabajo se pretende demostrar la inexistencia de la "falsa creencia" sobre que mejores valoraciones recibidas por los docentes son debidas a que estos otorgan unas notas más elevadas a sus alumnos/as, sin importar la dispersión de éstas. De esta forma, se pretende prevenir posibles comportamientos atípicos por parte de los docentes. De este modo, este trabajo arroja luz a la hora de (re)considerar qué variables deberían realmente tenerse en cuenta en las instituciones educativas, analizando sus procesos de enseñanza-aprendizaje, management y hasta la retención de sus actores clave.

Palabras clave. *metodologías educativas, motivación del estudiante, preferencias del estudiante, evaluaciones del profesor, optimización*

INTRODUCTION

The management of educational institutions is not easy at all as it brings together many actors such as educators, students, and managers, all of them essential when it comes to achieving one of the most complicated and important tasks in a country: the education of its future generations (Heffernan, 2022).

Imagine that you were the general director of a business school in Spain and you could choose for a moment each and every one of the members that would form part of the faculty. Which type of professors would you select? With what features? Would you opt for gender parity among them or would you favor one of them being more represented? Should everyone necessarily teach in English or would there be a percentage of classes in Spanish? Would they be hired full time or on the contrary would it be better to hire them part time so that they could have "real and updated" practical experience in different companies? Would you rather they were young or have a certain age and tenure? Now imagine you had retrospective data of such institution for the quintessential degree from any business school: '*Business Administration and Management'*, during four complete academic years (18/19, 19/20, 20/21, 21/22), which equals 210 courses, 1,320 European credits or ECTS[1] and 10,500 students, with all its grades, existing grade dispersion of the class, and evaluations made to their professors. Would you be able to draw conclusions to improve the functioning of this institution (or any other)? And specifically, would you be able to identify the courses and professors that motivate your students the most when it comes to obtaining better academic results (i.e., to obtain higher average grades)? And the courses and professors most valued by them (i.e., to the ones that they provide better evaluations)? Finally, could you clarify if there is any kind of relationship between both variables? In other words, would better academic results obtained by students in the different courses correspond to better evaluations provided to these professors? Or would there not be such a relationship?

To do so, the present work will begin by defining the individuals who receive this education nowadays, the so-called *'gen Z'* students, who were born between 1997 and 2012 (Choudhary & Pandita, 2023) and are considerably different from the students of past generations. And it is that, this type of students has the nothing despicable 25% less attention capacity than the generation immediately before them (the *'gen X'*), and, in addition, they verify 40% less the reliability or origin of the sources consulted (Google, 2019). That is to say, of every four concepts explained in the classroom, one of them will not be retained by the students (25%), and almost half will be assumed as something that "is so" and that "cannot be refuted" (40%), therefore decreasing the critical

[1] The European Credit Transfer and Accumulation System (ECTS) helps students and universities define and understand the workload that comes with lectures and study programs. Usually, each year of full-time study (or work, where applicable) is worth 60 ECTS. The conversion rate from ECTS to American credits is 2:1, meaning 60 ECTS would convert into 30 American credits. However, this conversion rate may vary, as some universities in the U.S. may use different credit systems (Ravshanbek, 2022).

capacity of the students (Villagrasa et al., 2022). In fact, this type of students has values, concerns, aspirations, way of learning, and even, a different brain structure, in this case due to their compulsive and constant use of mobile phones and screens (Bigné et al, 2019; Hutton et al., 2020; Jurayev, 2023), yet another proof, in this case tremendously objective, of previously mentioned differences. This situation has provoked that the techniques and methodologies currently used by professors, usually based on a more "traditional" or "classical" methodology (e.g., master lectures, repetition of concepts and memorization) are not fully effective, generating a palpable gap between the learning method used and the own students, often materialized through a lack of motivation, and therefore hindering their learning and attention capacity (Izagirre-Olaizola, et., 2020; Winquist & Carlson, 2014).

This is why, educational institutions should know and be concerned about the typology of professors they have in their faculty, as well as the type of methodology they use in class (or will probably use regarding its characteristics) since, surprisingly, they could have great effects on the student motivation to obtain better academic results; but also on their preferences for receiving this content only from some of them (i.e., even though the content taught would be the same, students would not be indifferent in the way it is given/by the one who gives it).

LITERATURE REVIEW

Consistent with previous literature, there is a clear positive link between student motivation and its academic performance (Ali et al., 2010; Huitt, 2001; Pinar-Pérez et al., 2021), understanding motivation as a state of mind that stimulates the activities and actions developed by the individuals (Ali et al., 2010). Increasing student motivation is not an easy task, however, its study has drawn the attention of a high number of scholars over time (Huitt, 2001; Morris et al., 2022; Ryan & Deci, 2020) with the aim of identifying the different factors that influence it. Among them, it will be of vital importance to decipher whether they generate what is known as intrinsic motivation or, on the contrary, increase extrinsic motivation. The latter refers when students get motivated to learn in order to achieve a mere objective, a reward, or avoid punishment, both external to the educational activity itself (Dev, 1997; Morris et al., 2022; Ryan & Deci, 2020; Silva et al., 2020). Conversely, an intrinsically motivated student would not need a reward or punishment to be engaged with the learning process itself and would find in it interest and enjoyment. Therefore, as Ali et al. (2010) argued:

> Students with intrinsic motivation will be more enthusiastic, self-driven, and proactive, being able to feel pleasure for their studies, and (however) students with extrinsic motivation will drag more when carrying out their academic tasks, will feel compelled to learn, and will always allocate minimal efforts to achieve maximum performances. (Ali et al., 2010, pp. 81).

Consequently, and as prior research shows, academic performance will be positively influenced by intrinsic motivation, whereas negatively affected by extrinsic motivation (Ali et al, 2010; Morris et al., 2022; Ryan & Deci, 2020). Accordingly, the main purpose of every educational institution should be to find professors that stimulate the intrinsic motivation of its students, thus achieving a better academic performance, as well as trying to get said students to value positively the work carried out by these professors in the teaching-learning process.

Academic performance and grade dispersion

At the same time and in a logical way, previous research shows a clear negative relationship between grade dispersion and academic performance since, as expected, less disperse classes would simply generate a general improvement in the learning process, therefore provoking higher

overall academic results in the students (Ferrer-Torregrosa et al., 2016; Sabot & Wakeman-Linn, 1991; Strello et al, 2021).

Likewise, literature also delves into this area distinguishing between *'soft skills' and 'hard skills'* when determining the prior relationship. *'Soft skills'* are all those skills that a priori seem intangible, such as the ability to work in a team, resilience, critical thinking, constant learning, digital skills or data-based decision-making (Hendarman & Cantner, 2018). For their part, *'hard skills'* represent technical or experiential knowledge that allows a worker to dominate a specific area or process within an organization (Hendarman & Cantner, 2018). In this way, literature particularly links high academic results (and therefore, low dispersion) with *'soft skills'* courses through what it calls as a *'grade inflation'* (Sabot & Wakeman-Linn, 1991); thus differentiating among departments such as *'Art', 'English', 'Business management', 'Philosophy', 'Psychology'*, and *'Political Science'* with a clear high-grading or '*grade inflation'* (and therefore, low dispersion), and departments such *as 'Economics', 'Chemistry', 'Physics'* and *'Mathematics'*, with low-grading (and therefore, high dispersion). The reasoning is quite straightforward, as '*soft skills'* courses:

- They are usually easier to understand and learn than *'hard skills'* courses, presenting a lower degree of difficulty (Hendarman & Cantner, 2018).
- They have a more practical and applied content to the real world and, therefore, are easier to memorize and do not forget (Wats & Wats, 2009).
- They are more vocational for *'Business Administration and Management'* students, meaning that these students would choose this degree due to their preferences towards such business and entrepreneurship courses rather than more technical ones. Therefore, as Chell and Athayde (2011) argue, vocationally oriented courses would allow students to more easily develop and improve the contents covered in said courses.

In addition, different studies similarly show empirical wide evidence on this matter as, for instance, the work developed by Castedo et al. (2017) with some *'engineering'* courses (*'hard skills'*), or Mateos-Ronco et al. (2011) with several *'accounting'* courses (*'hard skills'*), as well as Díaz-Silva (2018) with certain high school courses related to *'communication' ('soft skills')* and Manninen and Yli-Piipari (2021) with different *'physical education'* courses *('soft skills').* However, the common problem of these studies is that they all merely apply course-specific analysis (Castedo et al., 2017).

On the contrary, there are also other studies available in the literature that suffer from the opposite problem, being too generic with their analysis (Heffernan, 2022). Thus, for instance Beran and Violato's (2005) study of 370,000 courses and Centra's (2009) study of 238,000 courses focus their research across multiple universities and different countries. Nevertheless, ample evidence indicates that academic results (and also professors' evaluations) are strongly correlated with the teaching academic's demographics such as gender, ethnicity, language, perceived sexual identity, age or visible disabilities and other issues unrelated to the course (Uttl & Smibert, 2017). These other aspects, may even include classroom design, cleanliness of the university, quality of course websites, library services, food options available on campus, and difficulty in the admissions process for first year students (Heffernan, 2022). Therefore, despite wide studies would expand the replicability and representativeness of the research conclusions, they could be counterintuitive and not comparable. Hence, the most ideal option would be to analyze more than just one course or a group of them, but without mixing countries or educational institutions (Heffernan, 2022), and therefore, its demographics and particular characteristics. In fact, this would be the option followed in this paper, where only one educational institution will be under observation but where, however, consistent information about several years and absolutely all academic courses and professors would be carefully collected.

Potential effects of students' evaluations

As we have seen, that certain extraneous variables alone impact on academicresults and professors' evaluations demonstrates just how flawed the system of evaluations is, and to what extent results are determined by factors not related to course content or teaching quality (Rosen, 2018).

However, a starting point of many publications examining professors' evaluations is the acknowledgement of the assessment work made by the internal quality department of the university. That way, they put into value the sense that they provide the institution with data relating to the course design, its delivery, and the teaching staff performance (Heffernan, 2022). In fact, this "apparent data quality" is what generally leads institutions to use it as signifiers of teaching standards. Hence, many institutions expect their staff to achieve a certain result (for instance, a punctuation of four out of five or higher, or a mark over 80% of the total possible points, and so on) to be seen as fulfilling their duties. Stark and Freishtat (2014) also found that some universities even use evaluations to intentionally incite continuous cycles of competition amongst professors by making the acceptable result the one that is above the average. That is to say, no matter what the cohort's results are, half of the teaching staff would be above the average, but half of them would be below it and susceptible to the negative repercussions of not meeting the preset goals.

And this is a problem because compiling all the evaluations appears to provide a somehow objective picture of each professor/course success. However, what is rarely considered, or perhaps not seen, are the conditions and vicissitudes that shape the results, such as for instance, the specific demographic variables priorly mentioned or the existing differences between degrees, academic institutions, countries, or even type of courses (Stark & Freishtat, 2014). Thus, the inherent problem with these evaluations, is that they disguise such prejudices and biases, affecting subsequently the decisions made according to these data due to such assessments are also used to gauge professors' quality and are usually a key component in judging who is hired, who is let go, and who is promoted (Rosen, 2018; Stark & Freishtat, 2014).

Consequently, in this research we will try to take into account those issues by only analyzing one institution, and by creating different groups of variables comparable to each other, such as the type of course taught (more related to *'soft skills'* or *'hard skills'),* but also others based on the age of the professor, their more academic or practical profile, the vehicular language used in their classes, and some other variables subject to affecting teaching and easily accessible by the educational centers in order to make comparisons between courses and professors; and that at the same time move away from simple averages obtained in the whole sample, educational institution or degree, which are far removed from the reality experienced by each one individually (Uttl & Smibert, 2017).

Similarly, we will check if there are differences between such groups of variables when assessing professors' evaluations. In fact, some educators could be perfectly motivated to set easier assessments, or perhaps to grade more positively in order to facilitate the reception of better evaluations by their students (Heffernan, 2022). And this perception would be totally licit due to the priorly mentioned effects on their hiring, recognition, or promotion (Rosen, 2018; Stark & Freishtat, 2014). Because students evaluations do not really measure teaching effectiveness. Because students need to get good grades. And because faculty members also need to get good evaluations (Stroebe, 2020). Therefore, such evaluations could empower students to shape faculty behavior towards their own benefit. Nevertheless, prior literature shows that this relationship would only exist when students obtain a normal grade dispersion in class and thus each student gets the grade they deserve (Cagigal, 2016). That is, in general, students would prefer to obtain a differentiation from the worse students of the class, therefore avoiding situations where all of them obtain similar grades and there are no discernible differences.

OBJECTIVES

With the aim of solving the aforementioned problems, in this paper we propose to generate a sort of monitoring tool or automatic dashboard based on only one educational institution, with information about the academic results and grade dispersion obtained by the students in each course of the degree, as well as professors' evaluations. In particular, this tool would allow us to compare and consult the results obtained by different students and professors in individual courses, groups of similar courses, or even among both types of courses but in prior academic years, applying in each case the same or a different methodology, and in this way avoiding the potential problem arising from the concrete specificities that each individual course could need.

To be more concrete, this tool would benefit students, professors, and educational institutions as it would allow them to compare their results with other related data and not with generalist variables that are far away from their reality such as the aseptic average obtained in the whole university where they operate. In this way, in first place this tool would help the students to better understand the difficulty of one academic year versus another, due to the type and proportion of *'soft skills'* versus *'hard skills'* courses included in each academic course. Or due to the type of professors teaching those courses if for instance historically they gave higher or lower grades than the average. Something that could also help them to foresee and make better predictions about their chances to apply to some scholarships, Erasmus programs, master's degrees, awards, recognitions, etc., after knowing the possible evolution of their grades throughout the years (for example, between the 1st academic year and the last one, the average grade of the students is increased by more than one point: from 5.6 to 6.7 and the grade dispersion reduced by more than 10%: from 46% to 35%).

But, at the same time, this tool could also help educational institutions or professors to identify potential deviations from their "normal" or "expected" performance, both voluntarily or involuntarily. The first one would be produced by their inner motivation to secure their job, salary, or image through better evaluations. As a kind of false "glorification" or "support" obtained from their students. However, the second one would be produced due to their lack of knowledge about the level of demand of such institution, their own level of self-demand, or their previous experience carried forward in other educational centers. To be more specific, this deviation or gap would be easily discovered by comparing their individual course with the same course in previous years (for instance, *'course A'* in the academic year 22/23 with the exact same *'course A'* but in the academic year 23/24), with another one very similar to it (for instance, *'course B1'* taught in the 1st academic year with *'course B2'* taught in the 2nd academic year and considered as a continuation of the first one) or with the group of courses that it belongs to (for instance, the one of *'hard skills'*). Or even, realizing that a mere comparison of an individual course with the set of all the courses of a complete academic year (for instance, all the courses of the 1st, 2nd, 3rd, or the 4th academic year) or, worse still, the whole degree (i.e., the average of the four years) would not make any sense because they would contain and mix courses belonging to different categories and difficulties, and therefore the comparison will not be appropriate *per se*.

Different pathways to obtain the same output

As has just been mentioned, when determining the type of professors or *'dream team'* that an educational institution has or would like to have as part of its faculty, it would be extremely useful to count on reliable data and accumulated empirical evidence. Thus, knowing the academic performance of the students in each course, the existing grade dispersion, and the evaluations obtained by professors through such monitoring tool (let us briefly remember, that the ideal situation would consist of finding professors who managed to awaken the intrinsic motivation of the students,

generated low grade dispersion, and were valued positively), would greatly help to avoid basing decisions on mere subjective assumptions (Villagrasa et al., 2018; Villagrasa et al., 2024)..

In this way, students could foresee their future academic evolution. And for its part, contracted professors (or "chosen" to form part of the faculty team) could follow certain instructions or recommendations made by the educational institution based on its prior experience and data. However, it should be noted that of course, all professors would have their academic liberty, their methodology or, in other words, their way of doing things and giving their classes. So, could each professor stimulate the intrinsic motivation of the students in a different way in order to help them obtaining better overall results? The answer is yes. Of course. And this would be due to the aforementioned academic liberty of each professor, which together with its own expertise would help them to obtain the desired results.

Consequently, despite the existence of many types of methodologies to achieve student motivation and positive academic results, could we at least determine what type of categories or groups of courses could be more important in this sense (or even more valued/preferred by the students that effectively receive such teaching)? The answer in this case would also be affirmative. And, in fact, this data could even affect the type of professors hired since, for instance, professors who teach their classes in English (within the category: "taught in English" in Table 2) could be perfectly better paid/treated than those who teach in their mother tongue, due to the added difficulty involved in preparing their classes so that they looked interesting/motivating (and therefore the students obtained good academic results) and appreciated by the students (and therefore they valued these professors positively); or professors with a certain age or position in the industry (within the category: "*'boomer'* professor"[2] or "main activity: industry (more practical-oriented)" in Table 2), who could also receive higher compensation/treatment for their acquired experience or practical approach of their classes, which possibly would translate into a better teaching (transferred in higher academic results of the students) and perception on the part of the students (through their evaluations). Or at least, this is what logic tells us. However, we will see that it is not entirely true and that, thanks to the proposed tool, better and objective information could be obtained at the time to make more optimal decisions (Wilson, 2022).

METHODOLOGY

Sample

To examine the proposed relationships along this manuscript, we used a sample composed by data of the degree in *'Business Administration and Management'* of *'EDEM-Business School'*[3] (Valencia, Spain) during the academic years 18/19, 19/20, 20/21, 21/22. This dataset contained both objective and subjective data from 210 courses (which equals 1,320 ECTS) and 10,500 students, regarding its academic results, existing grade dispersion of the class, and evaluations made to their professors (see Table 1 below).

To get access to this sensitive information about students and professors, the authors of this paper signed the corresponding confidentiality agreements before gaining access to the data provided. Moreover, this information was proportioned in a totally anonymous and aggregated way. Likewise, to prevent the variability in the data collected from being attributable to the measurement

[2] *It is important to mention that according to literature, "boomers" are those individuals who were born between 1946 and 1964 (Choudhary & Pandita, 2023). This category is followed by the "generation X" or "gen X" who gathers the cohort born between 1965 and 1980. In turn, the latter group is followed by the "gen Y" or "millennials" born between 1981 and 1996, which is followed by the "gen Z", previously defined in the text as the ones born between 1997 and 2012 (Choudhary & Pandita, 2023).*

[3] *EDEM-Business School' is a Spanish academic institution founded in 2002. Currently, the school is affiliated with the two main public universities in the city: the 'Universidad of Valencia' and the 'Polytechnic University of Valencia'. In addition, the institution has an educational offer of 50 programs, including both university and management training, it houses a total of 2,500 students and employs more than 500 professors and researchers. Likewise, it is important to highlight that 'EDEM-Business School' is part of 'Marina de Empresas', a consolidated business ecosystem made up of a business accelerator, 'Lanzadera', an investment vehicle, 'Angels', and an educational center, 'EDEM-Business School', concentrated in what has become one of the most important and successful business hubs in the Mediterranean area.*

methods used or the presence of the own authors, we used the independent and anonymous evaluations made by the students to their professors through the internal quality system of the school. Therefore, any type of distortion by being subjectively measured by one individual is beyond question. Furthermore, in order to check the veracity of our analyses we tested for residuals behavior, linearity among variables and existence of collinearity between them. No significant problems were observed in any of the preceding categories (full analyses are available from the authors upon request).

In particular, the sample shows an average academic result of the students of 6.10 points4 (min. 3.40 – max. 8.70 out of 10), an average grade dispersion of 41% (min. 18% – max. 74% out of 100%), and an average of the evaluations made to their professors of 4.07 (min. 2.70 – max. 4.80 out of 5). At the same time, 35% of the respondents were female students, while 65% were male students. Besides, 56% of the courses were related to 'soft skills', while 44% to 'hard skills' ones (see Table 1 below).

Table 1. Description of the dataset.

MAIN INFORMATION GATHERED IN THE SAMPLE	
Degree	Business Administration and Management
Institution	EDEM-Business School
Country	Spain
Academic years	Four : 18/19, 19/20, 20/21, 21/22
Number of ECTS	1,320 ECTS
Number of courses	210 courses
Number of groups of courses	10 groups of courses
'Soft skills' vs *'hard skills'* courses	56% *'soft skills'* courses – 44% *'hard skills'* courses
Number of professors	240 professors
Number of students	10,500 students
Gender of students	65% male students – 35% female students
Academic performance of students	$\bar{x}$ = 6.10 *(min. = 3.40 – max. = 8.70)*
Grade dispersion of students	$\bar{x}$ = 41% *(min. = 18% – max. = 74%)*
Evaluations obtained by professors	$\bar{x}$ = 4.07 *(min. = 2.70 – max. = 4.80)*
'MVP professors' with evaluations > 4.5	24 *(only 2 professors reached this figure 4 times = 33%)*

In fact, it is important to note that, even though *'soft skills'* and *'hard skills'* are both present somehow within all the different courses (Hendarman & Cantner, 2018), each course will be always more related to one of them (and similarly will help students to develop them in a higher level) (Chell & Athayde, 2011). Finally, following the classification previously developed by Börner et al. (2018), in turn, all the courses belonging to *'soft skills'* and *'hard skills'* could be split in two, therefore creating four categorizations or groups of different courses: 1. *'business'* (46% of the courses of our sample) and 2. *'law'* (10% of the courses of our sample), linked with the *'soft skills'* (in total, 56%); and 3. *'economics'* (16% of the courses of our sample), and 4. *'technical'* (28% of the courses of our sample), related to the *'hard skills'* (in total, 44%).

[4] *It is important to highlight that the academic result of the student is an individual grade obtained exclusively by the student in the 'final exam' of the course, therefore, it could not be biased by the potential group work marks, collected in this case within the 'final grade' of the course.*

It was this classification and similar other research such as Heffernan (2022) or Uttl and Smibert (2017) which made us to also develop comparisons between other types of categories or groups of courses that could be of particular interest to our institution or any other, and that at the same time would be quite easily measurable and accessible by them. Besides, this partitioning would allow us to move away from simple comparison averages from the whole educational institution, the whole degree, or the whole academic year, which are far removed from the reality experienced by each course and its own demographics individually (Stark & Freishtat, 2014). Thus, for example, apart from analyzing whether there were differences between 1) *'soft skills'* and *'hard skills'* courses, we also assessed the potential divergence among 2) professors that taught in the 1st, 2nd, 3rd or 4th academic year, 3) male and female professors, 4) courses taught in Spanish or English, 5) internal and permanent professors or associate professors (understanding an internal and permanent professor as someone who is embedded to the university, with a permanent job in it; while the associate professor would have a temporary contractual relationship with the institution, renewing the contract yearly based on their needs or performance), 6) professors that their main professional activity is teaching and working as university professors, therefore being more academic-oriented or the ones that teaching is rather a hobby and just a complementary and extra professional activity, therefore being more practical-oriented (for instance, a manager of Deloitte that teaches only two classes per week, allocating the majority of his/her time to work as a business consultant)[5], 7) *'boomer', 'gen X' or 'gen Y'* professors, 8) courses taught by one single professor or by more than one, 9) semester courses or yearly ones, and 10) courses taught during the Covid-19 or no-Covid-19 period (understanding the Covid-19 period as the second semester of the academic year 19/20, where only 22 courses among all academic years were taught, or what is the same, they would represent only 11% of the 210 courses analyzed in our sample). For a better visualization of these categories and its representation within the total sample, see Table 2 next.

Differences among categories: which one really matters?

Following prior categories or groups of courses shown in Table 2, Table 3 expands this information by providing data on the academic performance of the students, the existing grade dispersion, and the evaluations obtained by professors in each of them. In this way, if within each of the ten categories (for instance, "academic year") there is a significant difference among its components (i.e., among the "1st academic year", "2nd academic year", "3rd academic year", and "4th academic year"), higher values are colored yellow (representing the sun or "the highest") and lower values are colored blue (representing the sea or "the lowest"). If they were not colored (i.e., there is a white color), that would mean that both values are close to each other. Next, within such colored components it is checked whether there is a significant difference among them and the average of the sample (for example, 6.10 in the case of academic results, 41% in the case of grade dispersion, and 4.07 in the case of the evaluations obtained by professors). If so, such yellow/blue colors are kept. On the contrary, they are blurred, coloring them in light yellow/blue.

[5] *In order to differentiate group 5 (internal and permanent professors vs associate professors) from group 6 (professors that their main professional activity is teaching and working as university professors vs the ones that teach as a hobby and just as a complementary and extra professional activity) we will provide the following examples. Thus, despite for internal and permanent professors (see group 5) the teaching activity would obviously represent their only and main professional task (see group 6), we could also find other different options. Such as associate professors that also work in different other universities part-time (see group 5), therefore, and as it happened with the first example, having a pure academic-oriented profile (see group 6). However, it could also be the case that those associate professors (see group 5) had their main work in the industry and taught only in their spare time, as a hobby, and just as a complementary and extra professional activity (see group 6), therefore applying a more practical-oriented approach in their classes (as opposed to the two previous examples).*

Table 2. Categories or groups of courses, and their representation within the total sample.

CATEGORIES OR GROUPS OF COURSES		% OF TOTAL	
All courses		**100%**	
1	Business courses (*'soft skills'*)	46%	**100%**
	Law courses (*'soft skills'*)	10%	
	Economics courses (*'hard skills'*)	16%	
	Technical courses (*'hard skills'*)	28%	
2	1st academic year	26%	**100%**
	2nd academic year	23%	
	3rd academic year	31%	
	4th academic year	20%	
3	Male professor	79%	**100%**
	Female professor	21%	
4	Taught in Spanish	42%	**100%**
	Taught in English	58%	
5	Internal and permanent professor	34%	**100%**
	Associate professor	66%	
6	Main activity: professor (more academic-oriented)	51%	**100%**
	Main activity: industry (more practical-oriented)	49%	
7	*'Boomer'* professor	19%	**100%**
	'Gen X' professor	45%	
	'Gen Y' professor	36%	
8	Taught by one professor	86%	**100%**
	Taught by more than one professor	14%	
9	Semester course	94%	**100%**
	Yearly course	6%	
10	Taught during no-Covid-19 period	89%	**100%**
	Taught during Covid-19 period	11%	

Table 3. Categories or groups of courses, together with their academic performance, grade dispersion, and evaluation obtained by professors.

	CATEGORIES OR GROUPS OF COURSES	Academic performance *(average grade)*	Grade dispersion *(top 3 vs rest of the class)*	Evaluation obtained by professors *(average grade)*
	All courses	**6.10**	**41%**	**4.07**
1	Business courses (*'soft skills'*)	6.50	35%	4.11
	Law courses (*'soft skills'*)	6.30	37%	4.15
	Economics courses (*'hard skills'*)	5.70	45%	4.10
	Technical courses (*'hard skills'*)	5.70	48%	3.97
2	1st academic year	5.60	46%	4.02
	2nd academic year	5.80	43%	4.02
	3rd academic year	6.50	36%	4.06
	4th academic year	6.70	35%	4.13
3	Male professor	6.10	40%	4.11
	Female professor	5.90	44%	4.00
4	Taught in Spanish	5.80	44%	4.18
	Taught in English	6.30	39%	3.98
5	Internal and permanent professor	5.70	46%	4.19
	Associate professor	6.20	39%	4.01
6	Main activity : professor (more academic-oriented)	5.90	43%	4.12
	Main activity : industry (more practical-oriented)	6.20	38%	4.01
7	*'Boomer'* professor	6.10	39%	3.93
	'Gen X' professor	6.10	42%	4.07
	'Gen Y' professor	5.90	42%	4.07
8	Taught by one professor	6.10	41%	4.06
	Taught by more than one professor	5.70	41%	4.07
9	Semester course	6.10	41%	4.06
	Yearly course	5.60	44%	4.20
10	Taught during no-Covid period	6.10	41%	4.10
	Taught during Covid period	6.30	38%	3.87

Calculation of grade dispersion

It is worth noting that this paper contemplates the existence of different ways to evaluate the grade dispersion within a classroom that are present in the literature (c.f., Castedo et al., 2017; Díaz-Silva, 2018; Mateos-Ronco et al., 2011 among others) such as the *'variance', 'standard deviation', 'coefficient of variation', 'boxplot'* or *'quartiles'*. However, to avoid diffusing the focus of this study, a detailed comparison between them is not carried out, directly selecting the measure that compares the 'grade obtained by the top students in the class with the rest of students'.

Nonetheless, in general, we could argue that no other article that uses, or simply shows, distinct dispersion measurements, delves into the meaning or way of calculating them, nor criticizes its difficulty to be interpreted, focusing only on providing them in a mere table and determining if there is more or less dispersion depending on whether the values are higher or lower compared to the rest of the samples/categories (e.g., Díaz-Silva, 2018; Mateos-Ronco et al., 2011, etc.). Simple as that. However, a mere comparison of the existing dispersion values between different samples/categories would not be enough for our study, where we would like to know the actual dispersion of the sample (with a concrete number), but besides being able to know if this dispersion could be considered as high or low with the naked eye (that is, with said concrete number but that, in addition, was easily interpretable and representative).

Thus, for instance, it would be difficult to know if a *'variance'* (which is calculated by taking the differences between each number in the data set and the average value), *'standard deviation'* (which is calculated as the square root of the *'variance'*) or *'coefficient of variation'* (which is calculated by dividing the *'standard deviation'* of the data set by the average value) of '0.5 points' or '1.5 points' could be considered as something relatively high or low, versus the ease of interpreting a +50% or a +70% of the 'grade of the top students of the class versus the rest of pupils' as a high dispersion in a sample –as put into practice by the research proposed in this paper–. Another example would be the difficulty of visually interpreting (and even more, being able to compare between a large number of samples –remember that our study amounts 210 courses–) the existing dispersion within a sample through the analysis of a simple rectangle provided by the *'boxplot'* (which through a more or less elongated rectangle gives a visual indication of how a data set is distributed and varies from its mean) or through the information of three mere individual data provided by the dispersion measure of the *'quartiles'* (which allow to determine the proximity or distance between the four equal groups into which a population can be divided); both of them, measures that would not provide a concrete individual number about the grade dispersion within a classroom as the previous ones did (*'variance', 'standard deviation',* or *'variation coefficient'*), but which, furthermore, would still not be easily interpretable (as was the case of the '*variance', 'standard deviation',* or *'coefficient of variation'*).

Likewise, note that due to the fact that the dispersion obtained by the 'top 3' students compared to the rest of the class, and the one obtained by the 'top 10' students compared to the rest of the class, keeps always the same proportion and a minimum difference of less than 5%, in order to facilitate the understanding of the different analyses carried out in the following sections of the work, from now on only the first one will be used –that is, the one of the 'top 3' students compared to the rest of the class–, taking into account that the second measure could also be perfectly used, without generating any type of variation in the logic and argumentation of the manuscript.

Finally, someone could have argued, for example, a modification in this calculation, using the 'top 20' of the class instead of the 'top 3'/'top 10'. However, the reason behind using 3/10 students or 20 students, would come determined by the sample size. Thus, in our case, where most of the classes have between 30 and 50 students, the use of 20 students to calculate the dispersion would eliminate the coherence of this calculation: to show the difference between the top students and the rest of the class.

RESULTS AND DISCUSSION

In the first place, observing Table 3, and in particular the first two columns related to "academic performance" and "grade dispersion", we could determine that none of the categories are significant regarding the average but two: 1. If they *are 'soft skills'/'hard skills'* courses and 2. If they belong to the 1st/2nd/3rd/4th academic year. In other words, none of the methodologies applied by different other professors in the rest of categories or groups of courses would affect the intrinsic motivation of the students and their academic performance, therefore not showing a significant level. However, this statement would not be entirely correct since, if we look closer, we will realize that the mere presence of *'soft skills'/'hard skills'* courses would be the unique factor affecting this relationship. And it is that, if we delve a little bit more into these results, we could identify that in the 1st and 2nd academic years there is only a 30% of *'soft skills'* courses versus a presence of 70% and 100% in the 3rd and 4th academic years. Consequently, the results would be higher (and therefore, the dispersion lower; Ferrer-Torregrosa et al., 2016) in the last courses simply due to the existence of more *'soft skills'* courses, which are much easier to understand, have more practical content, and are more vocational to '*Business Administration and Management'* students (Chell & Athayde, 2011; Hendarman & Cantner, 2018; Sabot & Wakeman-Linn, 1991; Wats & Wats, 2009).

In second place, observing the third column of Table 3 related to the "evaluation obtained by professors" we can similarly determine that there are significant differences between 1. *'Soft skills'/'hard skills'* courses, but also between 2. Courses taught in Spanish/English, 3. Courses with an internal and permanent professor/associate professor, 4. Courses with a *'boomer'/'gen X'/'gen Y'* professor, 5. Semester/yearly courses, and 6. Courses taught during Covid-19/no-Covid-19 period. From these results, it should be noted that if this information were known by professors, it could even affect their own perception about the future evaluations received (Heffernan, 2022). This, in turn, could make them voluntarily provide higher academic grades to their students in order to obtain better evaluations and feel happier, or even secure their salary and job. Especially if they belonged to a category where historically worse evaluations were received by the students, such as if they "taught their course in English", if they were "*'boomer'* professors", or if they worked such as "associate professors" (see Table 3). And this could be even more pronounced when said professors had junior profiles, started teaching a new course recently, or were new in the educational institution, due to their intrinsic desire to do well and be liked (Rosen, 2018; Stark & Freishtat, 2014). As a kind of false "glorification" or "support" obtained from their students and, therefore, from their university.

However, it is important to highlight that this difference in grades with respect to what is "normal" or "expected" could also occur involuntarily by the professor, due to their lack of knowledge about the level of demand of such institution, their own level of self-demand, or their previous experience carried forward in other educational centers.

In any case, nothing is further from reality, and although some professors could think that the grades given to their students voluntarily or involuntarily positively affect their evaluations, there is not such reciprocity. In fact, as shown by the literature, that reciprocity would only exist between the expectation that the student has about its grades (defined as the difference between the grade in the evaluated course and the average of its grades during all the degree) and the evaluation that he/she gives to the professor (Cagigal, 2016). Consequently, students would only positively evaluate professors in the presence of a normal grade dispersion in the class where each student obtains the grade they deserve (Cagigal, 2016). That is, the student would prefer the smart student to stand out from the rest, avoiding situations where all students get "surprisingly" good or bad grades, but all in unison.

In fact, if we check the data available in Table 3, the categories or groups of courses that provide higher grades than the average (>6.10), only in 33% of the cases obtain better evaluations than the average (>4.07). Or, put another way, in 66% of cases they would obtain worse evaluations.

Similarly, the categories or groups of courses that provide lower grades than the average (<6.10), only in 33% of the cases obtain lower evaluations than the average (<4.07), therefore following the precepts established by the literature. In this way, through this manuscript, we corroborate that higher grades, usually associated with lower grade dispersions (Ferrer-Torregrosa et al., 2016; Sabot & Wakeman-Linn, 1991; Strello et al, 2021), do not provide better evaluations by students (Cagigal, 2016). And we manage to demonstrate this in a study not focused solely on a course or a group of courses (Castedo et al., 2017), but on an entire educational institution (Heffernan, 2022), and over four academic years, thus contributing to the literature and its call when analyzing this aspect from a broad perspective but at the same time not subject to demographic or particular variations (Uttl & Smibert, 2017).

CONCLUSIONS

Prior research links elevated academic results with *'soft skills'* courses through what it calls as a *'grade inflation'* (Sabot & Wakeman-Linn, 1991). The reason is simple: the internal characteristics of this type of courses make students to excel (Chell & Athayde, 2011; Hendarman & Cantner, 2018; Wats & Wats, 2009). Nevertheless, the main criticism received by this type of studies is that they all merely apply course-specific analysis (Castedo et al., 2017). On the contrary, there are also other studies available in the literature that suffer from the opposite problem, being too generic with their analysis by analyzing thousands of courses from different universities around the globe (Heffernan, 2022). However, ample evidence indicates that academic results and professors' evaluations are strongly correlated with demographic variables, rarely considered by this research. Consequently, the conditions and vicissitudes that shape these results would be somehow disguised and mixed, hindering the comparison between courses or a group of them (Heffernan, 2022).

This study disentangles this problem by not only analyzing one course or a group of courses (Castedo et al., 2017), but an entire educational institution (Heffernan, 2022), and by creating different groups of variables comparable to each other (Uttl & Smibert, 2017); therefore, moving away from simple averages obtained in the whole sample, educational institution or degree, which are far removed from the reality experienced by each course and its own demographics individually (Stark & Freishtat, 2014).

In second place, this manuscript corroborates that higher grades, generally associated with lower grade dispersions (Ferrer-Torregrosa et al., 2016; Sabot & Wakeman-Linn, 1991; Strello et al, 2021), do not provide better evaluations by students (Cagigal, 2016). Again, these results were not assessed solely on a course or a group of courses (Castedo et al., 2017), but on an entire educational institution, during four complete academic years, and with comparable group of variables (Heffernan, 2022); widening the perspective of this study and not being subject to any demographic or particular restriction (Uttl & Smibert, 2017).

In this vein, this work aims to improve the pedagogy or "false belief" when it comes to explain the absence of any type of relationship between high grades and positive evaluations (Rosen, 2018; Stark & Freishtat, 2014). In fact, the study provides exemplary data on the "casually" worse evaluation given by students to the courses that provide better grades, as well as the better evaluation given to the courses that provide worse grades. Something that would break such "faith" and would clearly show that this relationship (as established in the literature; Cagigal, 2016) does not depend on any other factor than the existence of a normal grade dispersion in the classroom; which would allow "showing" the effective existing differences between the good and the bad students within a classroom.

The third conclusion of the study would come when determining the type of professors or *'dream team'* that a particular institution would like to have, maintain, or hire. Thus, looking at Table 3 we were able to affirm that the only typology or category that would affect the motivation of the students

and therefore, their academic performance, would be if they are *'soft skills'/'hard skills'* courses, due to their internal properties previously mentioned and explained in detail (Chell & Athayde, 2011; Hendarman & Cantner, 2018; Wats & Wats, 2009). Put differently, none of the methodologies applied by the professors of the rest of categories of courses would affect the intrinsic motivation of the students and their academic performance, therefore not showing a significant level in Table 3. Likewise, it is important to mention that the own distribution of *'soft skills'/'hard skills'* courses is determined by the degree curriculum itself and would not be modifiable by the school, as it is pre-established by the national educational plans. Therefore, we can conclude that regrettably academic institutions could not do anything to improve the overall grades obtained by their students.

However, it will be other categories the ones that would influence the preferences or evaluations of the students with the teaching received (and actually paid by them). In this case, again we would find whether the courses are *'soft skills'/'hard skills'* ones; but also, whether they are taught in Spanish/English, have an internal and permanent/associate professor, have a *'boomer'/'gen X'/'gen Y'* professor, are semester/yearly courses, or were taught during Covid-19/non-Covid-19 period. Consequently, due to the relevance and valuation by the students of these aspects, it is recommended that they were properly taken into consideration when ranking, rewarding, hiring, or even retaining the different professors by the institution. And it is that, this type of professors are the ones that the students, throughout the different academic years, value best, considering them as a key piece of their training when acquiring the same knowledge (or academic performance), although perhaps, explained in a certain way they particularly like or with a specific methodology. Therefore, we can conclude that educational institutions should try to identify such *'renowned professors'* or *'MVP professors'* due to their great relevance and weight in the entity. With the aim of deepening in this matter, we carried out a descriptive analysis within the 240 professors that formed part of our sample. Among all of them, we discovered that only in twenty-four of the times any professor obtained an evaluation higher than 4.5 out of 5, and curiously, only two professors reached this score in four times (each one). Consequently representing 33% of the total. That is why, it would be recommendable that this institution in particular takes the corresponding actions in order to keep these professors as part of its faculty, and secondly, perhaps transfers this information to them so that they feel more motivated and aware of their importance within the institution they work for.

Finally, this study proportions a handy monitoring tool through which it would be gathered trustworthy information and accumulated empirical evidence about the academic performance, grade dispersion and professors' evaluations of an academic institution. In this way it would be possible to deeply help students, professors, and management teams of such educational institutions to better place them and know their current individual performance in the aforementioned variables. Besides, this information could also help these actors to make better future decisions based on objective and contrasted data, as opposed to mere subjections or beliefs (Villagrasa et al., 2018; Villagrasa et al., 2024). In this vein, the creation of different groups of variables comparable to each other and strongly accessible (Uttl & Smibert, 2017) would make a difference, moving away from the current simple average comparisons with unrelated data such as the mean obtained by the whole educational institution, the whole degree, or the whole academic year (Stark & Freishtat, 2014), where there would be a mix of courses belonging to different categories and difficulties, and where the comparison would not be appropriate *per se*.

As examples of its applicability, we could mention that this tool would profoundly help students to know the historical results obtained by other pupils in the past, as well as determining the potential future evolution of their grades along the different academic years to apply to different scholarships, masters' programmes, or awards. In fact, following Table 3, students could easily understand that the only category which would affect the evolution of their grades would be the number of *'soft skills'* or *'hard skills'* courses present during each academic year. And as previously

observed, the percentage of such *'soft skills'* courses increases from 30% in the 1st academic year until 100% in the last one, showing in this way the improvement that students could obtain on average.

On the side of professors and educational institutions, the effect of this tool would also be extremely effective. Thus, faculty members could receive certain instructions or recommendations based on the historical results and experience gathered by the institution. Or maybe, they could simply make their own consultations. The goal would be that even following their own academic liberty and methodology to achieve the higher intrinsic motivation and academic performance in the students as possible, professors could identify potential unintentional deviations from their "standard" or "expected" performance. Especially if they did not have prior consolidated teaching experience or felt a little bit insecure with their job. Consequently, if for instance a particular professor was teaching the course *'Mathematics II'* in the 2nd academic year he/she could consult or compare his/her results with the same course in previous years (for instance, *'Mathematics II'* in the academic year 22/23 with *'Mathematics II'* in 23/24), with similar individual courses (i.e., related courses taught in prior academic years with the same students, such as *'Basic statistics'* and *'Statistical inference'* in the case of *'Econometrics'; or 'Mathematics I'* in the case of *'Mathematics II'*), or with the complete group of courses they belong to (for instance, *'hard skills'*) (Börner et al., 2018; Chell & Athayde, 2011; Hendarman & Cantner, 2018). Or even identify the lack of adequacy when comparing an individual course with the set of all the courses of a complete academic year (for example, all the courses of the 1st, 2nd, 3rd or 4th academic year), the whole degree (that is to say, the average of the four academic years), or, worse still, all the degrees taught in their school (something that, although may seem an aberration, often times is provided by the internal quality system of the educational institutions), as they would be mixing courses that belong to different categories, difficulties and nature (Heffernan, 2022; Uttl & Smibert, 2017). In addition, as previously argued, with this tool it would not be necessary to know the particular methodology applied by the professor in the classroom, thus avoiding the problem derived from the concrete specificities of each course or the open access to this information protected by the data protection law.

Limitations and future research avenues

Like any study, ours has limitations that can set the stage for future research opportunities. First, following prior literature (Heffernan, 2022) we intentionally focused on only one degree and one country, the one in *'Business Administration and Management'* taught in 'EDEM-Business School' (Valencia, Spain) where despite existing a great parity between *'soft skills'* and *'hard skills'* courses (56% vs 44% according to Table 1), it is true that their distribution throughout the different academic years varies, with the percentage of *'soft skills'* courses being much higher in the last two years. Therefore, future research should compare these results with other degrees or countries where the distribution of *'soft skills'/'hard skills'* courses (or any other unbalanced variable in our sample, such as the age of the professors, their more academic or practical profile, the language in which classes are taught, etc.) was more even along the academic years (or along the degree itself). Likewise, this study could also be carried out in some degrees where it was known in advanced that some courses such as the *'soft skills'/'hard skills'* ones were more predominant (c.f. an Engineering or Architecture degree).

Besides, despite belonging to the group of *'soft skills'* or *'hard skills'* courses (or any other category), it could be the case that some professors did motivate the students, but not in the right way, for example, working on the extrinsic motivation of the pupils instead of on the intrinsic one. Consequently, and as research shows, academic performance would be negatively affected by such type of motivation (Ali et al, 2010). As can be imagined, this situation could be easily detected when comparing such course with their peers (for instance, with the complete group of '*soft skills'/'hard skills'* courses). Thus, future research should delve into the goals and strategies for these professors which potentially would be to transfer the often short-term thinking and 'not very

motivating' of the traditional, master and repetitive lectures (related with the extrinsic motivation, and where an outcome or reward is presented in terms of the student final grade; Dev, 1997), to the long term, motivation and enthusiasm for the teaching received (in this case, more related with the intrinsic motivation, and where the student could finally have a good fit with the teaching method proposed, challenging itself, stimulating its autonomous learning, and therefore enabling its self-education and interests; Dev, 1997).

However, it could be also possible that some professors were motivating the students in the right way, and that the "actual problem" was that their courses belonged to a group with historically low grades on average. For instance, the *'hard skills'* ones. In fact, if we remember, this occurs thanks to the *'grade inflation'* (Sabot & Wakeman-Linn, 1991) which makes the *'hard skills'* courses to have lower grades and higher dispersion, while the *'soft skills'* courses behave in just the opposite way, having higher grades and lower dispersion, due to they are easier to understand (Hendarman & Cantner, 2018), with more practical content (Wats & Wats, 2009), and more vocational to *'Business Administration and Management'* students (Chell & Athayde, 2011). Consequently, even though these professors worked hard to improve the students' motivation, the own nature of this type of courses would not make possible a great improvement in their academic results. Or at least until one preset threshold. Nevertheless, future research could investigate about the possibility of creating complementary activities out of the official sessions of the degree, where combining *'hard skills'* courses with *'soft skills'* ones in order to link their contents, but also to take advantage of the "benefits" that *'soft skills'* courses could create over the *'hard skills'* ones (due to the aforementioned characteristics of these variables), making them more accessible, understandable, and close.

Finally, to provide a good and accurate feedback, as well as an appropriate forecast of the variables analyzed in this study, it is important that they were updated frequently, which means that these variables should be gathered, grouped, and calculated yearly. However, in spite of this, the conclusions obtained could be somehow biased due to aspects not related to the demographics of the students (Uttl & Smibert, 2017) but to different others, such as the day on which the exam was taken with respect to the rest of the courses (i.e., it would not be the same to take the exam of a *'hard skills'* course the first day of the examination period than the last one, where potential worse results could be generated) or even if the student had taken the exam for the first time, for the second time or for the fourth time. The same reasoning could be applied for the day when professors' evaluations were made. Thus, for instance, some students could provide better evaluations if the day of the assessment the professor decided to develop a more dynamic, interactive, and relaxed activity as opposed to a pure theoretical lecture. Or even whether the professor caught their attention for talking in the classroom or set an unscheduled exam, something that could be certainly avoided if the professor knew when the evaluation was going to be developed. Future research should delve into this matter taking into account how *'gen Z'* university students, the current recipients of the teaching, really are, behave, interact and socialize among each other's, therefore affecting the evolution of these results (Fernández-Moya et al., 2020; Hutton et al., 2020).

REFERENCES

Ali, I., Hamid, K., & Khan, M. A. (2010). A study of university students' motivation and its relationship with their academic performance. *International Journal of Business and Management*, *5*(4), 80-88. https://doi.org/10.2139/ssrn.2899435

Beran, T., & Violato, C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter?. *Assessment & Evaluation in Higher Education*, *30*(6), 593-601. https://doi.org/10.1080/02602930500260688

Bigné, E., Badenes-Rocha, A., Ruiz, C., & Andreu, L. (2019). Development of a blended course for online teaching: process and outcomes. *Journal of Management and Business Education*, *2*(2), 108-126. https://doi.org/10.35564/jmbe.2019.0010

Börner, K., Scrivner, O., Gallant, M., Ma, S., Liu, X., Chewning, K., ... & Evans, J. A. (2018). Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy. *Proceedings of the National Academy of Sciences*, *115*(50), 12630-12637. https://doi.org/10.1073/pnas.1804247115

Cagigal, J. L. (2016). Relaciones entre notas del alumno y evaluaciones al profesor. *Revistapuce*, *102*, 169-198. https://doi.org/10.26807/revpuce.v0i102.10

Castedo, R., López, L. M., Ortega, M. F., Cabrera, J. D., García-Martínez, M. J., Sanchidrián, J. A., ... & Paredes, C. (2017). Aula invertida para la mejora del aprendizaje en la asignatura de Transferencia de Calor y Materia. In *IV Congreso Internacional sobre Aprendizaje, Innovación y Competitividad (CINAIC 2017)*, pp. 231-236. https://doi.org/10.26754/CINAIC.2017.000001_048

Centra, J. A. (2009). Differences in responses to the student instructional report: Is it bias. *Princeton, NJ: Educational Testing Service*. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=4145fb9109ac090cd280e0bb59d24b266c602071

Chell, E., & Athayde, R. (2011). Planning for uncertainty: soft skills, hard skills and innovation. *Reflective Practice*, *12*(5), 615-628. https://doi.org/10.1080/14623943.2011.601561

Choudhary, H., & Pandita, D. (2023). Maximizing learning outcomes in the digital age: the role of microlearning for Gen Z. *Development and Learning in Organizations: An International Journal*. https://doi.org/10.1108/DLO-02-2023-0038

Dev, P. C. (1997). Intrinsic motivation and academic achievement: What does their relationship imply for the classroom teacher? *Remedial and Special Education*, *18*(1), 12-19. https://doi.org/10.1177/074193259701800104

Díaz-Silva, L. A. (2018). *Diseño y aplicación de una estrategia de lectura "Procedimientos Lógicos Básicos" para mejorar la comprensión lectora de textos escritos en el Área de Comunicación del tercer grado de educación secundaria de la Institución Educativa El Cumbe-Provincia de Cutervo*. Universidad nacional de educación Enrique Guzmán y Valle. Retrieved from https://repositorio.une.edu.pe/handle/20.500.14039/3429 [accessed on 25 April 2023]

Fernández-Moya, M., Cuadros, P. J., Salvador, C., & Pinar J. M. (2020). The Montessori method in university teaching. In *INTED2020 Proceedings*, pp. 3861-3864. https://doi.org/10.21125/inted.2020.1079

Ferrer-Torregrosa, J., Jiménez-Rodríguez, M. Á., Torralba-Estelles, J., Garzón-Farinós, F., Pérez-Bermejo, M., & Fernández-Ehrling, N. (2016). Distance learning ECTS and flipped classroom in the anatomy learning: comparative study of the use of augmented reality, video and notes. *BMC Medical Education*, *16*(1), 1-9. https://doi.org/10.1186/s12909-016-0757-3

Google (2019). *Generation Y and Z: characteristics and differences.* Retrieved from https://www.thinkwithgoogle.com/intl/es-es/insights/tendencias-de-consumo/generaciones-y-y-generacion-z-en-que-se-diferencian-y-como-captar-su-atencion/ [accessed on 5 August 2023]

Heffernan, T. (2022). Sexism, racism, prejudice, and bias: A literature review and synthesis of research surrounding student evaluations of courses and teaching. *Assessment & Evaluation in Higher Education*, *47*(1), 144-154. https://doi.org/10.1080/02602938.2021.1888075

Hendarman, A. F., & Cantner, U. (2018). Soft skills, hard skills, and individual innovativeness. *Eurasian Business Review*, *8*(2), 139-169. https://doi.org/10.1007/s40821-017-0076-6

Huitt, W. (2001). *Motivation to learn: An overview. Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. Retrieved from http://www.edpsycinteractive.org/topics/motivation/motivate.html [accessed on 14 May 2023]

Hutton, J. S., Dudley, J., Horowitz-Kraus, T., DeWitt, T., & Holland, S. K. (2020). Associations between screen-based media use and brain white matter integrity in preschool-aged children. *JAMA Pediatrics*, *174*(1), e193869-e193869. https://doi.org/10.1001/jamapediatrics.2019.3869

Izagirre-Olaizola, J., Morandeira-Arca, J., Mitxeo-Grajirena, J., Mendizabal-Zubeldia, A., & Lertxundi-Lertxundi, A. (2020). Reinforcing the involvement of students in learning business economics through active methodologies and student response systems. *Journal of Management and Business Education*, *3*(1), 29-46. https://doi.org/10.35564/jmbe.2020.0004

Jurayev, T. N. (2023). The use of mobile learning applications in higher education institutes. *Advances in Mobile Learning Educational Research*, *3*(1), 610-620. https://doi.org/10.25082/AMLER.2023.01.010

Manninen, M., & Yli-Piipari, S. (2021). Ten practical strategies to motivate students in physical education: Psychological need-support approach. *Strategies*, *34*(2), 24-30. https://doi.org/10.1080/08924562.2020.1867270

Mateos-Ronco, A., Lajara-Camilleri, N., & Marín Sánchez, M. (2011). *Sistema de evaluación en contabilidad y resultados*. Retrieved from http://web.ua.es/es/ice/jornadas-redes-2011/documentos/posters/185422.pdf [accessed on 29 July 2023]

Montessori, M. (1912). *The Montessori Method*. New York, NY: Frederick A. Stokes Company. Retrieved from https://books.google.es/books/about/The_Montessori_Method.html?id=4F4WAAAAIAAJ&redir_esc=y [accessed on 10 May 2023]

Morris, L. S., Grehl, M. M., Rutter, S. B., Mehta, M., & Westwater, M. L. (2022). On what motivates us: a detailed review of intrinsic v. extrinsic motivation. *Psychological medicine*, 1-16. https://doi.org/10.1017/S0033291722001611

Pinar-Pérez, J. M., Morales-Arsenal, R., Fernandez-Moya, M., Cuadros-Solas, P., & Salvador, C. (2021). Mitigating deficiencies of generation Z through new educational methodologies in a business statistic course. In *Proceedings INNODOCT/20. International Conference on Innovation, Documentation and Education*, pp. 81-88. https://doi.org/10.4995/INN2020.2020.11821

Ravshanbek, J. (2022). Credit-module system, its basic principles and features. *Yosh Tadqiqotchi Jurnali*, *1*(4), 304-309. https://doi.org/10.5281/zenodo.6582173

Rosen, A. S. (2018). Correlations, trends and potential biases among publicly accessible web-based student evaluations of teaching: a large-scale study of RateMyProfessors. com data. *Assessment & Evaluation in Higher Education*, *43*(1), 31-44. https://doi.org/10.1080/02602938.2016.1276155

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary educational psychology*, *61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Sabot, R., & Wakeman-Linn, J. (1991). Grade inflation and course choice. *Journal of Economic Perspectives*, *5*(1), 159-170. https://doi.org/10.1257/jep.5.1.159

Silva, R., Rodrigues, R., & Leal, C. (2020). Student learning motivations in the field of management with (and without) gamification. *Journal of Management and Business Education*, *3*(1), 47-71. https://doi.org/10.35564/jmbe.2020.0005

Stark, P., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1

Strello, A., Strietholt, R., Steinmann, I., & Siepmann, C. (2021). Early tracking and different types of inequalities in achievement: Difference-in-differences evidence from 20 years of large-scale assessments. *Educational Assessment, Evaluation and Accountability*, *33*, 139-167. https://doi.org/10.1007/s11092-020-09346-4

Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology*, *42*(4), 276-294. https://doi.org/10.1080/01973533.2020.1756817

Uttl, B., & Smibert, D. (2017). Student evaluations of teaching: teaching quantitative courses can be hazardous to one's career. *PeerJ*, *5*, e3299. https://doi.org/10.7717/peerj.3299

Villagrasa, J., Buyl, T., & Escribá-Esteve, A. (2018). CEO satisfaction and intended strategic changes: The moderating role of performance cues. *Long Range Planning*, *51*(6), 894-910. https://doi.org/10.1016/j.lrp.2017.12.002Villagrasa, J., Donaldson, C., & López, C. (2022). Twitter as a tool to monitor attitudes: The strategic usage of social media. *Journal of Management and Business Education*, *5*(4), 392-423. https://doi.org/10.35564/jmbe.2022.0023

Villagrasa, J., Escribá-Esteve, A., & Donaldson, C. (2024). How can complacency mold managerial decisions? The role of perceptions in strategic decision-making. *Corporate Governance and Organizational Behavior Review*, 8(2), 212–231. https://doi.org/10.22495/cgobrv8i2p21

Wats, M., & Wats, R. K. (2009). Developing soft skills in students. *International Journal of Learning*, *15*(12). https://doi.org/10.18848/1447-9494/CGP/V15I12/46032

Wilson, S. (2022). Student perception of learning versus performance in project management education. *Journal of Management and Business Education*, *5*(1), 48-62. https://doi.org/10.35564/jmbe.2022.0004

Winquist, J. R., & Carlson, K. A. (2014). Flipped statistics class results: Better performance than lecture over one year later. *Journal of Statistics Education*, *22*(3), 1-10. https://doi.org/10.1080/10691898.2014.11889717