Online-Journals.org

iJEP International Journal of Engineering Pedagogy

iJEP | eISSN: 2192-4880 | Vol. 15 No. 6 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijep.v15i6.54591

PAPER

Gender Differences in Mathematics Achievement among Engineering Students: Does the Way Performance is Measured Matter?

Jose Luis Arroyo-Barrigüete (✉), Susana Carabias-López, Antonio Obregón, Yolanda González-Arechavala, Santiago Canales-Cano

Universidad Pontificia Comillas, Madrid, Spain

jlarroyo@comillas.edu

ABSTRACT

Recent research on the gender gap in mathematics achievement [1] has found no differences among Spanish undergraduate students in business administration degrees. This study aims to replicate the aforementioned work in an engineering school, which differs notably in its sample composition: a lower percentage of female students and a higher proportion of top-performing students in mathematics. Combining regression models with NeuralSens, a state-of-the-art algorithm based on interpretable neural networks, we analyze the academic achievement in two first-year mathematics courses (Algebra and Calculus) and one second-year course (Differential Equations), considering a sample of 1,832 undergraduate engineering students. NeuralSens is employed to verify that the linear regression specification captures the underlying relationships and that no relevant nonlinear effects have been omitted. Overall, female students perform as well as, or slightly better than, their male peers across the three courses, although the effect sizes are small. These results hold even in a context traditionally considered unfavorable to female students. Our findings highlight the importance of using comprehensive and continuous evaluation methods over isolated standardized tests when assessing mathematics achievement and suggest that female students' performance in engineering programs is not inferior when proper assessment methods are employed.

KEYWORDS

engineering, gender differences, academic achievement, mathematics, performance measurement

1 INTRODUCTION

The underrepresentation of women in science, technology, engineering, and mathematics (STEM) fields is a well-documented fact [2–6]. The Global Gender Gap Report states that "women are underrepresented in STEM fields, and the gender

Arroyo-Barrigüete, J. L., Carabias-López, S., Obregón, A., González-Arechavala, Y., Canales-Cano, S. (2025). Gender Differences in Mathematics Achievement among Engineering Students: Does the Way Performance is Measured Matter? *International Journal of Engineering Pedagogy (iJEP)*, 15(6), pp. 84–110. https://doi.org/10.3991/ijep.v15i6.54591

Article submitted 2025-01-23. Revision uploaded 2025-06-02. Final acceptance 2025-06-17.

gap is most prevalent in two fields: Information and Communication Technologies and Engineering and Manufacturing" [7, p. 42]. Academic literature reflects on the undesirable consequences of this situation. On the one hand, there is evidence that a greater representation of women in studies positively affects academic performance in both men and women [8, 9]. On the other hand, diversity in the workforce contributes to creativity, productivity, innovation, and success. As Corbett and Hill pointed out [10, p. 92], "women's experiences—along with men's experiences should inform and guide the direction of engineering and technical innovation."

The possible causes of this low female presence have been analyzed in multiple papers on STEM studies [11–19], particularly engineering [20, 21]. One of the causes of the gender gap in STEM fields seems to be the selection bias derived from the perception of a gender gap in terms of performance, that is, women's performance in STEM is lower than men's. Academic literature points out the relevance of this stereotype in women's decision-making [10, 22–27]. Stereotypes about women in STEM also generate structural biases in education [28, 29], and actions aimed at increasing the representation of minorities in STEM fields are ineffective if they do not consider inclusion measures that counteract these kinds of biases [30]. Multiple examples of inclusion initiatives for women in STEM are shown in the literature [20, 31–35].

When complex relationships between female representation and performance in STEM are studied, mathematics courses play a critical role. First, because individuals with unproductive mathematics dispositions are less likely to enroll in STEM studies [36]. Secondly, the association between sparse representation of women in STEM and low female achievement seems particularly strong in mathematics courses, which have a critical role in most STEM disciplines [8]. Growing literature supports that, in general, women's performance in mathematics is not lower than that of men. Meta-analyses conducted on youth populations [37, 38] and on both youth and adult populations [39] concluded that, on average, there is no gender gap in mathematics performance. However, this result should be qualified. On the one hand, there is a large cross-cultural variability: there are important differences between different countries in the direction and intensity of the gender gap in mathematics achievement [13, 33, 37, 40]. On the other hand, among high performers, differences seem to persist [41, 42], as observed, for example, in the Mathematics Olympiads [43], although it is true that the gap seems to be narrowing over time [33]. Finally, gender differences could vary depending on the kind of assessment: standardized tests, which are commonly considered objective measures, tend to show a male advantage, which is not observed in academic grades [44–48]. Sonnert and Fox [49, p. 98] concluded that "[t]he analysis of the relationship between standardized scores and grades among science/engineering majors and degree recipients in the same departments and institutions would further reveal dynamics of gender and academic performance among undergraduates in these critical fields."

The issue of a gender gap in women's performance in STEM, if it exists, can be interpreted from the perspective of fairness in assessment. As noted in the standards for educational and psychological testing [50. p. 51]. "Regardless of the purpose of testing, the goal of fairness is to maximize, to the extent possible, the opportunity for test takers to demonstrate their standing on the construct(s) the test is intended to measure." Fairness concerns major requirements of assessment, ranging from the validity of individual score interpretation to moral issues [50–53]. According to the American Educational Research Association [50], when a difference in the test functioning (DTF) for different specially defined groups occurs, potential sources of bias in measurement should be investigated. If construct-irrelevant variables are identified, DTF constitutes a predictive bias that would reveal a lack of fairness.

The first construct that literature highlights to explain the differences in performance in mathematics by gender is self-efficacy [54–61]. Perceived self-efficacy (SE) is a belief, which, according to Bandura [62], "is concerned with judgments of how well one can execute courses of action required to deal with prospective situations." Evidence that females have on average lower mathematics SE than males is shown in the meta-analysis of Else-Quest et al. [37], subsequent literature [63–65], and the systematic review of Wang and Yu [66]. According to the meta-analysis of Huang [67], this gap increases with age. Growing literature accepts this explanation for the gender gap in mathematics performance in standardized tests, favoring male learners [65, 68, 69], and difficulties in changing it because it is based on negative stereotypes of girls in mathematics [27, 55, 57, 58, 70]. However, this gender gap in SE seems to disappear for content and strategies that are directly related to what they have been taught [69, 71, 72]. This fact is coherent with previous research reporting the disappearance of the gender gap in performance, favoring men when achievement is measured by grades or by certain kinds of items in standardized tests. These findings have been documented in studies conducted across different countries [46, 73–77].

A second relevant explanation is related to math anxiety (MA), which is understood as anxiety triggered by the presence, memory, or anticipation of situations related to mathematics [78, 79]. According to social cognitive theory, one of the leading causes of anxiety arousal in a situational context is low perceived SE in exercising control over the related situational events. In fact, similar states with different interpreted causes could be felt and labeled as another emotion, such as anger [80]. There is empirical evidence of the association between low SE and MA [81–85]. Since females tend to have lower SE, gender differences could also be expected in MA. Indeed, literature reveals that female learners have a higher prevalence of MA [86–89], and the difference with males increases with the mathematics gender stereotype endorsement [90–91].

According to different meta-analyses, MA and math achievement tend to have a negative association [54, 87, 89, 92–95]. With regard to gender mediation, according to the recent meta-analysis by [92] and [87], for a certain level of MA, the effect on achievement does not change between male and female students. The decrease in performance attributed to MA does not reflect a lower mathematics competence [96, 97], which means there is a predictable bias. The cognitive mechanism that causes this drop seems to be a reduction in available working memory capacity, which is particularly necessary in complex mathematical problems when MA states arise [96–101]. MA is especially significant under conditions of time pressure and high-stakes evaluation [102–108], which is frequently the case for standardized tests. It is in light of this evidence that Cipora et al. [78] propose to evaluate as a possible criterion for MA diagnosis "the discrepancy between math performance in relaxed, e.g., self-paced, and stressful, e.g., time pressure, high stakes situations" (p. 14). From this perspective, the MA construct gives great importance to the study of this discrepancy in the gender gap in mathematics performance as one of the implicit measures of MA that can be considered more reliable measures than questionnaires [109].

Most of the studies mentioned focused on the relationship between gender and academic performance, particularly among science/engineering majors, and were carried out in US universities. It is therefore important to extend these works to other geographical contexts in order to assess the extent to which similar results are obtained that could be interpreted as an MA anxiety signal, according to Cipora et al. [78], and bring to light issues of fairness.

The present paper examines the performance in mathematics of engineering undergraduate students at a Spanish University, Universidad Pontificia Comillas, replicating the research developed by [1] in the area of business administration (BA). The previous study focused on first-year BA courses, Mathematics I and II, which covered topics in Algebra and Calculus. However, the academic context differs substantially: the level of mathematical difficulty in a BA is lower, and the proportion of female students is much higher than in engineering. Despite these differences, the interest lies in comparing relative mathematics performance under different assessment methods (standardized testing versus continuous evaluation), as both studies do. This replication is particularly valuable because it examines whether the results previously found are confirmed in a more demanding context, characterized by a lower percentage of female students and a higher proportion of top-performing mathematics students. The conclusions achieved by these researchers were that at the pre-university level, male students outperform their female peers (higher grades). Still, throughout the first year of university studies, there are no differences in grades, and in those cases in which differences exist, female students outperform their male peers. Among other possible explanations, the authors suggested that this result is due to how mathematics performance is assessed in each case: standardized tests for pre-university performance and grades for performance in undergraduate subjects.

This study works with the grades obtained in the subjects of Algebra and Calculus, both in the first year, and Differential Equations, in the second year of engineering degrees (industrial and telecommunications technologies degrees). The final grade depends on various types of assessment tests that are conducted, all of which are focused on problem-solving. An important issue that helps prevent instructor bias is that, in all three mathematics courses analyzed, both midterms and final exams are identical across all class groups. Additionally, each exam question is graded by a single designated instructor across all sections, ensuring a uniform grading criterion. This centralized grading system minimizes instructor-related variability and guarantees a high level of consistency across course sections. All exam accommodations and modifications recommended by a specialized team at the university for students with special needs are conveniently implemented.

From the point of view of fairness, interactions between assessment and teaching style are also relevant, provided that instructors are responsible for both the delivery of instruction and the assessment processes [50, 110, 111]. These interactions are a natural concern for faculty when designing the assessment system [110, 112]. For the case of the three selected subjects, two objectives are established for the assessment system to support learning: The first one is to provide ongoing feedback to students through their teachers, which is an important factor that encourages persistence in attaining an engineering degree, according to Amelink and Meszaros [113] and Khodadad [114]. The second objective is to incentivize all students towards consistently working in order to gradually acquire the knowledge and problem-solving skills they are going to be tested on, with a constructivism-inspired approach [115]. An active learning style is pursued by working on problem-solving skills in different contexts, in which students' level of independence is modulated. The possibility of consulting the teacher is always present. There is evidence that supportive and active learning have an inclusive effect in STEM [114, 116, 117]. Instructors' phronesis has managed to come very close to the recommendations to operationalize fairness in the assessment of a subject [53, 110]. However, the context incorporates two very relevant differences with respect to the replicated study in the area of business administration [1], which could foster the presence of a gender gap favoring men, as previously mentioned. The first is the low representation of women.

For engineering studies at a Spanish University, a low representation of women is shown [118]. In the sample considered in this paper, the percentage of women is close to 25%, while in [1], the percentage of women was close to (and in several cases exceeded) 50%. The second difference is related to the student profile. In the Spanish university system, engineering degrees are among the most demanding, so there is an important self-selection bias: only students with a higher level in sciences (especially physics and mathematics) opt for these degrees. This leads to a high concentration of top performers in these disciplines in engineering degrees. Given the previously mentioned evidence that there seems to be a gender gap in performance for this group, the present research also offers the opportunity to evaluate whether these differences by gender are indeed observed.

Thus, our research question is whether there is a gender gap in mathematics achievement favoring male students among undergraduate engineering students. The hypothesis to be tested is that no such gender gap exists when mathematics performance is measured through grades based on various formative activities rather than on standardized tests.

2 MATERIAL AND METHODS

2.1 Participants and procedure

The study was developed at the Universidad X, a medium-sized private Spanish university founded in 1890. The participants were 1,832 undergraduate engineering students (483 female and 1,349 male) who entered this university between the academic years 2015–2016 and 2019–2020 to study industrial or telecommunications technology degrees. The data were provided by the Office of Data Governance and Intelligence ("Oficina de Gobierno e Inteligencia del Dato") at Universidad Pontificia Comillas, under the authority of the Vice-Rectorate for Research, Faculty Affairs, and Artificial Intelligence. All records were fully anonymized to prevent the identification of individual students, in accordance with the university's ethical and legal data protection requirements. During the preparation of this work, the authors used ChatGPT to improve the manuscript's readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication. A quantitative non-experimental methodology has been chosen. The design has been structured in four phases: identification of the most relevant variables related to academic performance in mathematics, collection of information, crude data analysis, and statistical analysis through Neural Networks and linear regression models (OLS).

2.2 Variables

The dependent variables in the regression models are the grades obtained in the Algebra and Calculus subjects, both in the first year, and Differential Equations in the second year. These three courses were selected because they are the only mathematics subjects included in the engineering curriculum during the first two academic years. Other courses offered during this period cover topics such as physics, chemistry, technical drawing, programming, and introductory engineering, but none involve additional mathematics instruction. Focusing on these three subjects allows us to comprehensively capture students' mathematics performance during the early stage of their degree. All grades correspond to the first call. Because some students dropped out of school or did not take a particular subject, each regression includes a different number of individuals (refer to Table 1).

Table 1. Number of students and mean (SD) grades by subject and gender. Grades on a 0–10 Scale

Gender	N (Access)	N (Algebra)	Mean (SD) Algebra	N (Calculus)	Mean (SD) Calculus	N (DiffEq)	Mean (SD) DiffEq
Female	483	472	5.19 (2.32)	472	5.43 (2.25)	369	7.30 (1.87)
Male	1349	1321	4.68 (2.54)	1320	4.86 (2.45)	977	6.78 (1.95)
Total	1832	1793	4.81 (2.49)	1792	5.01 (2.41)	1346	6.92 (1.94)

As previously mentioned, the assessment systems employed in the three subjects fulfill two main objectives: firstly, providing continuous feedback from teachers to students and secondly, encouraging students to engage progressively in acquiring the knowledge and competencies required by each course. The final grade includes short problem-solving exams during regular class periods that last around one hour and count for 10% of the final grade on average; midterm exams, also focused on problem-solving, count for 25% of the final grade; computer software exams, in which students are asked to solve problems using software they have learned to use during practical classes, count for 5% of the final grade; and final problem-solving exams assess students' overall knowledge of the subject and are worth 60% of the final grade.

Regarding independent variables, a considerable number of causal factors affecting academic performance have been identified in previous research, and a recent meta-analysis on this topic [119] identified 105 different variables grouped in two major groups: factors related to the student and factors associated with the instructional process. Nevertheless, regarding the student-related factors, it is usual to distinguish between personal and social variables [120]. In this case study, both instructional and social factors play a minor role due to the homogeneity of both the students and the teaching process. On the one hand, students in Universidad Pontificia Comillas have a similar social background, because almost all students fall into the upper-middle class. A paper analyzing the business administration degree of the same university [121] concluded that 86% of fathers and 77% of mothers had a higher education degree. On the other hand, instructional factors are the same for the three subjects analyzed, because both teaching methodologies and evaluation systems are almost identical for all groups, without relevant differences among them.

So, this paper will focus on personal factors, with pre-university performance being the first of them. This factor is highly relevant since it combines the student's skills, work capacity, and compromise. Several papers have reported significant correlation with academic performance in university students [120, 122, 123]. As a proxy of pre-university performance, we have used the grade obtained in EvAU (Evaluación para el Acceso a la Universidad—Evaluation for University Access), equivalent to the SAT in the Spanish university system, which is used by all public universities and some private universities to select their future students. The EvAU, which ranges from 0 to 10, aggregates the grade from an exam and the grade from the last two years of high school. The regression models incorporate this variable and its square, as previous research has identified a possible non-linear effect [122]. We have also included grades in Universidad X's own admission exams, which include tests in Mathematics (AE Mathematics), Physics (AE Physics), English (AE English), and Abstract Reasoning (AE Reasoning). These four tests are carried out within a three-hour timeframe. Specifically, the mathematics test is very stressful for the candidates: it consists of 15 highly challenging multiple-choice questions that must be solved within 60 minutes.

Another relevant factor is where the students come from, as previous research points to an adverse effect due to the adaptation to a new life situation when a student changes his or her residence [124, 125]. This is an interesting peculiarity of the Spanish university system, and recent research has proved that this variable is highly relevant [122]. Thus, we have included a dummy variable (Madrid) to indicate whether students completed their previous studies in Madrid, the province where the university is located. This variable takes the value 1 for students coming from Madrid and 0 for the rest. Finally, we have incorporated the variable Female, which takes the value 1 for female and 0 for male students.

2.3 Data analysis

The R programming environment (R Core Team, 2025) has been used to develop statistical analysis. The model's estimation has been carried out by using basic functions included in such programming environments [126] and the "lmtest" [127], "ggplot2" [128], "fmsb" [129], "RCurl" [130], "car" [131], "WRS2" [132], "caret" [133], "dplyr" [134], "lfe" [135], "rstatix" [136], and "NeuralSens" [137] packages.

In the first stage, a crude analysis of the math level at the entrance to the degree (AE Mathematics), as well as the grade in EvAU, was developed. We have also developed a crude analysis for grades in three subjects: algebra, calculus, and differential equations. In all cases, a contrast of normality (Shapiro-Wilk) and a contrast of equality of variances (Levene's test) have been carried out. When normality is not verified, but equality of variances is, the Mann-Whitney-Wilcoxon test has been used. The Yuen test has been applied in those cases where both conditions are violated.

In the second stage, linear regression models have been fitted for the three subjects. We have standardized all variables to allow comparability of coefficients. No evidence of multicollinearity was found, as all variance inflation factors (VIFs) were below 5 across all models. On the other hand, we detected the existence of heteroscedasticity, which generates inference problems. Therefore, multivariate linear regression models with heteroskedasticity-consistent standard errors have been used. Residuals were inspected visually (Q-Q plots) and with formal tests (Shapiro-Wilk). Although slight deviations from normality were observed in some models, the relatively large sample size ensures that the validity of statistical inference is not compromised. Finally, we have verified these results by adjusting a neural network (NN) and employing a state-of-the-art algorithm, "NeuralSens" [137], that allows estimating sensitivities to input variables in an NN model. NeuralSens computes the partial derivatives of the NN output with respect to each input variable, providing a local sensitivity analysis that captures both linear and nonlinear relationships. Thus, if the relationship between an input variable and the outcome is truly linear, the mean sensitivity estimated by NeuralSens will be virtually identical to the beta coefficient obtained from an OLS regression. Conversely, if the relationship is nonlinear, the mean sensitivity will differ from the OLS coefficient, and this nonlinearity will be flagged by a high standard deviation in the distribution of local sensitivities. As a consequence, the model will automatically capture any nonlinearity since NNs do not require a priori functional specification. Therefore, in the case where the linear regression models are well specified, the NN results will be identical. In the case of discrepancies, we would have an indication that the regression model should be modified. In summary, the use of NN with post-hoc analysis in this study serves as a validation tool to ensure that the functional specification of the OLS models is appropriate. Unlike other machine learning methods, NeuralSens provides case-specific insights into how each input contributes to the output, making it particularly suited for assessing potential nonlinearities and confirming the robustness of the linear specification. The grid method has been employed to identify the optimal NN in each case, considering from 1 to 10 neurons in the hidden layer and decay from 10^{-7} to 10^{-2}. The criterion for selecting the best fit was the RMSE (root-mean-square error). To prevent overfitting problems, in all cases, 10-fold cross-validation has been used.

3 RESULTS

3.1 Crude analysis

Table 2 shows the scores obtained by male and female students in the different tests considered. It is relevant to mention the large difference in absolute value between the EvAU score (overall pre-university performance) and the score on the mathematics entrance test (pre-university performance in mathematics, measured using a test developed by Universidad Pontificia Comillas). The reasons for this discrepancy are the high demands of the admission process at Universidad Pontificia Comillas, whose entrance exams have a very high level of complexity, and the fact that the EvAU includes other subjects whose grades are usually higher than those obtained in mathematics. The low dispersion in EvAU scores can be attributed to the high selectivity of the admission process: only students with strong academic records are admitted, which naturally results in relatively low variability.

The crude analysis of the grades obtained in the entrance assessments, both in terms of the mathematics exam and the overall EvAU grade, shows that although females have a higher grade in the EvAU (significant differences at 95%), their performance in the university's mathematics exam is lower than that of male students (refer to Table 2, which reports p-values corrected for multiple comparisons using the Holm method). When the grades in the three subjects under study are analyzed, there are significant differences at 95% in all three cases in favor of female students. That is, women show a better performance both in their pre-university global achievement (measured by the EvAU score) and in undergraduate mathematics subjects. However, with regard to their pre-university performance in the specific case of mathematics (measured according to the entrance test), their performance is worse than that of their male classmates. Although statistically significant differences were observed, effect sizes were small in all cases except for the EvAU score. Importantly, in undergraduate mathematics subjects, the differences were in favor of female students.

Table 2. Statistical comparison of academic performance: Grades in the test of access in mathematics, EvAU, and the three subjects under analysis (algebra, calculus, and differential equations)

Gender	Access (Math)	Access (EvAU)	Algebra	Calculus	Dif. Equations
Female: mean (SD)	4.90 (2.00)	8.58 (0.91)	5.19 (2.32)	5.43 (2.25)	7.30 (1.87)
Male: mean (SD)	5.13 (1.96)	8.14 (0.96)	4.68 (2.54)	4.86 (2.45)	6.78 (1.95)
Contrast (p-value)	W = 304816 (0.035)	Y = 9.58 (<0.01)	Y = 3.61 (<0.01)	Y = 4.55 (<0.01)	W = 209780 (<0.01)
effect-size	0.05 (small)	0.36 (medium)	0.13 (small)	0.18 (small)	0.13 (small)

The histograms of grades are shown in Figure 1. Figure 2 presents box and whisker plots of grades in Algebra, Calculus, and Differential Equations, disaggregated by gender. Each box represents the interquartile range (IQR), with the horizontal black line indicating the median. Dotted horizontal lines correspond to group means (blue for female students and red for male students), highlighting slight differences in central tendency. Grade dispersion is greater in Algebra and Calculus than in Differential Equations for both genders. This is likely due to the fact that Differential Equations is a second-year course and, therefore, does not include students who dropped out of the degree program during the first year. As a result of this selection process, only the best students remain in the sample, leading to lower variability in performance.

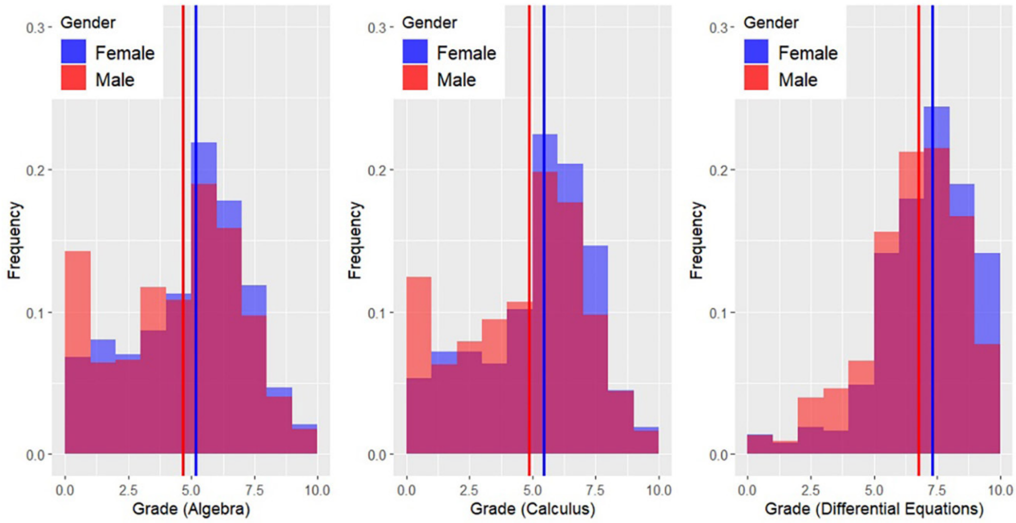

Fig. 1. Histogram for grades in Algebra, Calculus and Differential Equations (female vs. male students)

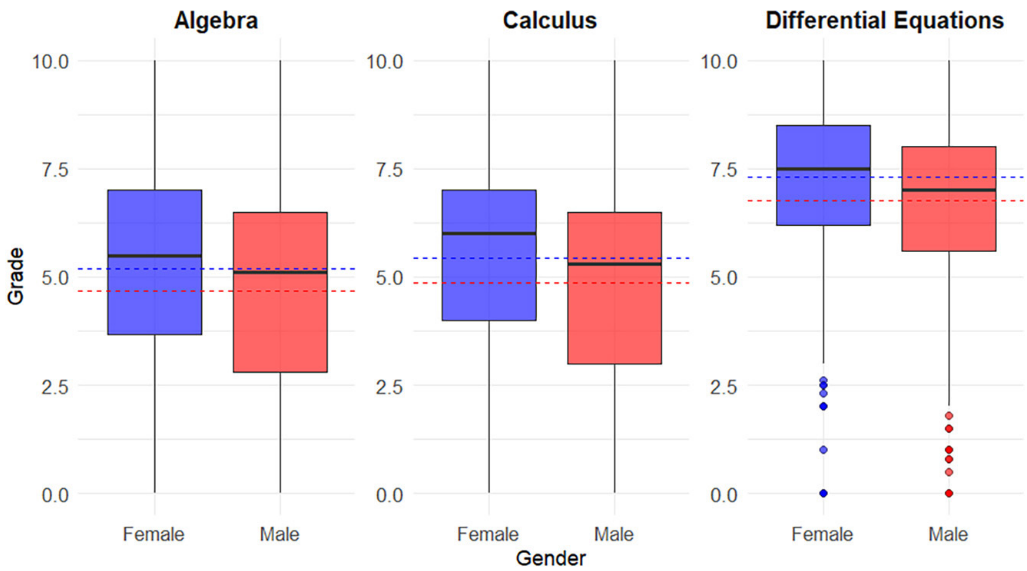

Fig. 2. Box and whisker plots for grades in Algebra, Calculus and Differential Equations (female vs. male students)

Given that in the Spanish university system the criterion for passing the course is to obtain a grade equal to or higher than 5 (on a 0–10 scale), the risk ratio has been calculated for the three subjects using this grade (refer to Table 3). It can be seen that the risk of failing is significantly lower for female students, while in all cases the risk ratio is lower than 1.

Table 3. Risk ratio for female students vs. male students in algebra, calculus, and differential equations (cut-off criterion = 5)

	Algebra	Calculus	Dif. Equations
Risk Ratio	0.78	0.72	0.54
Confidence Interval (95%)	0.67–0.90	0.61–0.85	0.35–0.83
p-value	< 0.01	< 0.01	< 0.01

3.2 Regression and neural network models

Tables 4 to 6 show the results of the regression models for the three subjects analyzed. It can be seen that in all of them, the grade obtained in EvAU and its quadratic term are significant. In fact, the EvAU variable is the one that generates the greatest effect. The grades obtained in the entrance exams in Mathematics and Physics are also significant in the three subjects. We have also confirmed the importance of the variable "Madrid," since not only is it significant in the three subjects, but it is the second most important variable after the EvAU grade in Algebra and Differential Equations.

Finally, in relation to gender, the variable under study in this paper, we do not observe differences in the case of Calculus (refer to Table 5), but differences exist in the subjects of Algebra (refer to Table 4) and Differential Equations (refer to Table 6). In these subjects, female students achieve better results than their male peers, although the magnitude of the effect is small. In other words, the mathematical gap identified at the university entrance (measured on the university entrance exam) disappears and is even reversed in two subjects. An analysis of dropout patterns shows that the proportion of female students among those who did not complete the course is very similar to that among those who did, across all three subjects. This suggests that attrition bias due to gender is likely minimal in this study and confirms that there is no significant selection bias that could compromise the interpretation of the results.

Table 4. Standardized OLS regression results for algebra

Variables	Algebra				
	Estimate	95% IC		p-Value	Effect Size
(Intercept)	–0.25	–0.33	–0.16	< 0.01	
AE Mathematics	0.14	0.10	0.18	< 0.01	Small
AE Physics	0.20	0.16	0.24	< 0.01	Small to moderate
AE English	–0.03	–0.08	0.01	0.10	
AE Reasoning	0.06	0.02	0.09	< 0.01	Negligible
EvAU	0.51	0.47	0.56	< 0.01	Large
EvAU^2	0.05	0.02	0.07	< 0.01	
Madrid	0.25	0.17	0.33	< 0.01	Moderate
Female	0.09	0.01	0.17	0.02	Small
R^2/Adjusted R^2	0.42/0.42				

Table 5. Standardized OLS regression results for calculus

Variables	Calculus				
	Estimate	95% IC		p-Value	Effect Size
(Intercept)	–0.17	–0.25	–0.09	< 0.01	
AE Mathematics	0.17	0.13	0.21	< 0.01	Small
AE Physics	0.12	0.09	0.16	< 0.01	Small
AE English	–0.07	–0.11	–0.03	< 0.01	Small
AE Reasoning	–0.01	–0.05	0.03	0.70	
EvAU	0.55	0.51	0.59	< 0.01	Large
EvAU^2	0.04	0.01	0.07	< 0.01	
Madrid	0.16	0.08	0.24	< 0.01	Small
Female	0.07	–0.01	0.15	0.09	
R^2/Adjusted R^2	0.40/0.40				

Table 6. Standardized OLS regression results for differential equations

Variables	Dif. Equations				
	Estimate	95% IC		p-Value	Effect Size
(Intercept)	–0.26	–0.37	–0.14	< 0.01	
AE Mathematics	0.13	0.07	0.19	< 0.01	Small
AE Physics	0.09	0.03	0.14	< 0.01	Small
AE English	–0.02	–0.08	0.03	0.38	
AE Reasoning	–0.04	–0.09	0.01	0.14	
EvAU	0.38	0.32	0.43	< 0.01	Moderate
EvAU^2	0.07	0.02	0.11	< 0.01	
Madrid	0.21	0.09	0.32	< 0.01	Small to moderate
Female	0.15	0.04	0.26	0.01	Small
R^2/Adjusted R^2	0.20/0.19				

Table 7 shows the results of the NN models for the three subjects considered, used to verify the results obtained in the regression models. In all cases, the optimal model resulted in having 1 neuron in the hidden layer (7-1-1), with the decay being slightly different in each subject, as shown in Table 7. Following Pizarroso et al. [137], if both the mean (mean sensitivity, MS) and standard deviation (sensitivity standard deviation, SSD) are near zero, it indicates that the variable is irrelevant. If the mean is different from zero and the standard deviation is near zero, it indicates that the variable is relevant and has a linear effect. If the standard deviation is different from zero, regardless of the value of the mean, the output has a non-linear relationship with the input, and as a consequence, the variable is relevant and has a non-linear effect. Finally, mean squared sensitivity (MSS) is a measure of the relevance of each variable.

As can be seen, the results obtained are virtually identical to those of the regression models. This confirms the absence of non-linear effects not included in these models. In fact, the NN points to a small nonlinearity in the EvAU variable (SSD higher than in the other variables), which is consistent with the functional specification of the regression models and with the results obtained: in all cases, the quadratic term was significant (refer to Tables 4–6). In conclusion, the results are remarkably robust, both for the sample size and the complete coincidence of results when two very different tools are applied. That is, the similarity between the regression coefficients and the mean sensitivities obtained from the NN provides additional support for the adequacy of the linear specification. The Appendix includes the graphical results obtained with NeuralSens for the three subjects (see Figures A1 to A3). These figures are relevant in that they allow us to verify that the sensitivity distributions (lower part of the figures) do not present abnormal distributions, such as more than one mode, which would be an indicator of some nonlinear effect.

Table 7. Neural network models for algebra, calculus, and differential equations

Variables	Algebra (7-1-1/decay = 10^{-6})			Calculus (7-1-1/decay = 10^{-4})			Dif. Equations (7-1-1/decay = 10^{-5})		
	MS	SSD	MSS	MS	SSD	MSS	MS	SSD	MSS
AE Mathematics	0.14	0.02	0.14	0.17	0.02	0.17	0.11	0.05	0.12
AE Physics	0.20	0.03	0.20	0.12	0.02	0.12	0.08	0.04	0.09
AE English	–0.04	0.01	0.04	–0.07	0.01	0.07	–0.05	0.02	0.06
AE Reasoning	0.06	0.01	0.06	–0.01	0.00	0.01	–0.01	0.00	0.01
EvAU	0.50	0.07	0.51	0.54	0.07	0.55	0.40	0.18	0.44
Madrid	0.24	0.03	0.25	0.16	0.02	0.16	0.20	0.09	0.22
Female	0.10	0.01	0.10	0.07	0.01	0.07	0.16	0.07	0.18
R^2	0.42			0.40			0.20		

Note: In each case, the network architecture and the selected decay are indicated.

Table 8 shows the results of the regression models for average grade obtained in the three subjects considered (Algebra, Calculus, and Differential Equations) for the 1,345 students who have taken all three courses. The table also includes the results of the equivalent NN model (average grade obtained in the three subjects considered as the dependent variable). Again, the results obtained are virtually identical to the regression model. The conclusions are very similar to those obtained in the individual regressions. Once again, the EvAU grade is the variable with the greatest impact, followed by "Madrid," and the grades obtained in the mathematics and physics admission exams. Gender is significant, indicating that female students have a higher average performance in mathematical subjects than their male counterparts, although the effect size is small.

Table 8. Standardized OLS regression results and NN for the average value in the three subjects (algebra, calculus, and differential equations)

Variables	OLS					7-1-1 Network Decay = 1e-06		
	Estimate	95% IC		p-Value	Effect Size	MS	SSD	MSS
(Intercept)	–0.30	–0.41	–0.20	< 0.01				
AE Mathematics	0.18	0.13	0.23	< 0.01	Small	0.16	0.06	0.17
AE Physics	0.18	0.14	0.23	< 0.01	Small	0.16	0.06	0.17
AE English	–0.06	–0.11	–0.02	0.01	Negligible	–0.08	0.03	0.08
AE Reasoning	0.02	–0.03	0.06	0.39		0.03	0.01	0.03
EvAU	0.51	0.46	0.56	< 0.01	Large	0.51	0.20	0.55
EvAU^2	0.11	0.07	0.14	< 0.01				
Madrid	0.22	0.12	0.32	< 0.01	Small to moderate	0.22	0.08	0.23
Female	0.14	0.05	0.23	< 0.01	Small	0.14	0.05	0.15
R^2	**0.39**					**0.39**		

4 DISCUSSION

The results of the regression models for the three subjects considered (Algebra, Calculus, and Differential Equations) show that the EvAU variable has the greatest effect, while the grades obtained in the entrance exams in Mathematics and Physics are also significant for all three subjects. This confirms that pre-university performance is highly relevant when predicting academic performance in undergraduate students [120–122, 138]. The importance of the variable "Madrid" is also noted. As we indicated previously, although this result has been identified in previous studies [122, 124, 125], it is still puzzling. In other countries, it is very common for university students to move away from their family home. This is much less common in Spain (approximately 30% in our sample), and we observe that this change has a relevant and negative impact on academic performance.

In relation to gender, the results obtained show a contrast between the performance in admission exams (AE) designed by the university and subsequent performance in undergraduate mathematics courses. Although female students score lower than their male counterparts in AE, which could corroborate previous studies on pre-university gender gaps, they perform equally well or slightly better in Algebra, Calculus, and Differential Equations. The gender variable is statistically significant in Algebra and Differential Equations but not in Calculus, with consistently small effect sizes favoring female students. While subject-specific differences may reflect distinct cognitive demands, exploring these variations lies beyond the scope of this study.

This discrepancy in outcomes may be attributable to the nature of evaluation methods. As mentioned, the AE corresponds to standard tests with multiple-choice questions. They are solved within a short period, under high pressure for students, and significantly impact their academic life. These conditions are prone to generating anxiety, among other reasons, due to their competitive nature, in which men tend to perform better than women [139], partly due to the adverse environment created [43, 140]. This observation is also consistent with the mentioned research on undergraduate studies in business administration [1]: as highlighted by Rozgonjuk et al. [141], STEM students have a higher perceived self-efficacy than social science students, but the prevalence of MA is similar. On the other hand, when the evaluation of performance considers a larger number of elements, demonstrating a more continuous and persevering effort over time, and is not limited to the result of a single standard test, the results of the female students tend to be better than those of the male students. This is the case with the grades in mathematics subjects analyzed in this study

The outputs obtained from the NN models for the three subjects are virtually identical to those of the regression models. This finding indicates that our results are highly robust, supported by both the sample size and the consistency of outcomes derived from applying two methodologically distinct approaches. Indeed, the similarity between the regression coefficients and the mean sensitivities obtained from the NN provides additional support for the adequacy of the linear specification.

Taken together, these results confirm our hypothesis: there is no gender gap favoring male students when mathematics performance is measured through grades based on various formative activities rather than on standardized tests. This conclusion is consistent with previous research among undergraduate students in general [77] and among undergraduate Spanish students in other disciplines [1]. Nevertheless, these results raise several interesting points. As previously indicated, our sample is characterized by a low percentage of women and a high percentage of top-performing students, both factors that could potentially lead to a gap favoring men. Yet, this is not the case. Hence, even in a situation we might consider unfavorable to female students, their performance is not inferior in any way to that of their male peers. The gap observed in AE suggests that the method used to measure academic performance (standardized tests versus classroom grades) impacts outcomes, as previously pointed out in existing literature [45, 46, 77, 106]. However, because our data are observational and the analytical approach is non-experimental, all reported relationships must be viewed as correlational. Future experimental studies will be necessary to establish definitive causal links.

5 CONCLUSIONS

In accordance with the purpose of this article, the main conclusion that can be drawn is that no relevant differences in mathematics performance by gender have been observed at the level of Spanish undergraduate engineering students, or, if there were, they would be in favor of female students. This statement refers to situations where the measurement of mathematical performance encompasses the evaluation of various formative activities that demonstrate a significant number of students' capabilities, rather than solely relying on standardized tests conducted under conditions of time pressure and high-stakes assessment.

The absence of a gender gap in the academic achievement of female and male students in mathematics at the higher educational level coincides with the result obtained in recent research among undergraduate students in business administration at the same University [1]. It is important to note that, although the previous study focused on Algebra and Calculus, the present study also includes Differential Equations, and the mathematical content in the engineering program is more advanced. Furthermore, the gender composition of the student body is notably different. The fact that no gender gap in engineering studies is observed acquires an important significance, because of some fundamental differences in the circumstances in which both studies are carried out. On the one hand, it is necessary to consider that the relationship between the demand and supply of places in Spain to study engineering leads to the fact that, in general, only highly qualified students access these degrees. These students are high performers, especially in the disciplines of physics and mathematics, a group of students where, as has been indicated, literature continues to perceive the persistence of a certain gender gap. On the other hand, it should be noted that the low percentage of female students in engineering degrees, in an academic context of fewer female professors and in a professional context of relevant positions occupied mainly by men, can affect their self-efficacy perception. Given these elements, which could lead to an unfavorable situation for female students' academic performance, it is important to verify that this performance is not worse but, if anything, better.

The plurality of tools used methodologically reinforces the conclusion obtained in this work and the robustness of the results observed. Both the use of regression models and NN models, which are applied in this paper, lead to consistent results: the gender variable, if significant, is significant in confirming a higher performance of female students in mathematical subjects.

6 IMPLICATIONS FOR ENGINEERING EDUCATION AND FURTHER RESEARCH

The previous conclusion highlights the importance of considering potential biases introduced by assessment methodologies, particularly standardized tests, when interpreting gender differences in mathematics performance. As indicated by Ashcraft and Moore [99, p. 204], "Math-anxious individuals avoid elective math coursework, avoid college majors that require math, and avoid career paths that involve math." A significant body of literature suggests that MA is more prevalent among women. Although our study does not directly measure math anxiety, the observed differences between performance in standardized admission tests and grades obtained through continuous assessment might suggest the influence of factors such as anxiety or test stress. Future research explicitly measuring MA could help clarify the extent to which anxiety contributes to the observed gender differences in standardized math assessments.

It is worth acknowledging that standardized tests are sometimes chosen as a measurement formula due to time constraints or the need to use homogeneous and seemingly objective instruments for a large number of individuals. Nevertheless, it is important to be aware that performance on these tests might be influenced by external factors, such as anxiety due to stressful testing conditions. Standardized tests are also chosen because of the ease of implementing them online, which could be one of the causes of the increase in the gender gap among engineering students in favor of men during the COVID-19 lockdown [142].

As Hopko et al. [97, p. 663] noted, different evaluation methods lead to different results, which "elucidates the importance of clearly defining skills to be measured and the corresponding strategy of measurement." Our findings appear to align with this insight, as the observed differences between standardized admission tests and continuous formative assessments suggest that evaluation methods may influence measured academic outcomes. As a consequence, to promote equity in mathematics achievement, instructors should balance high-stakes examinations with formative assessments, which are observed to have an inclusive effect in STEM.

Such a blend (a) mitigates the performance volatility induced by test stress, a factor shown to penalize many female students, and (b) increases the amount of feedback students receive, a practice consistently linked to persistence and retention in engineering programs.

Our conclusions are limited to a specific educational system, such as the Spanish one. So, they allow us to assume that the absence of a gender gap in university performance in mathematics-related subjects extends beyond the Anglo-Saxon sphere, the orbit to which most of the similar studies carried out to date refer. The main limitation of this study is that the sample consists of students from a single university with highly homogeneous evaluation practices. As a result, the findings may not be fully generalized to institutions with different grading cultures or student populations. It must be added that only the degrees in industrial and telecommunications technologies have been analyzed. Nevertheless, these two engineering degrees are the most common in Spain and, in addition, are included in the areas with the highest gender gap, as noted above [7]. The consistency of the results, obtained using different models, and their substantial coincidence with the previous study at a business school, confirms their soundness, at least within the defined territorial framework. However, it would be interesting to replicate this work in other universities so that our conclusions may or may not be confirmed.

7 ACKNOWLEDGMENTS

The authors acknowledge the funding of this work by the Santalucía Chair of Analytics for Education.

We are grateful to the Office of Data Governance and Intelligence ("Oficina de Gobierno e Inteligencia del Dato") at Universidad Pontificia Comillas for their invaluable support in providing access to the anonymized data used in this study.

8 REFERENCES

[1] J. L. Arroyo-Barrigüete, S. Carabias-López, F. Borrás-Pala, and G. Martín-Antón, "Gender differences in mathematics achievement: The case of a business school in Spain," *SAGE Open*, vol. 13, no. 2, pp. 1–14, 2023. https://doi.org/10.1177/21582440231166922

[2] H. Blackburn, "The status of women in STEM in higher education: A review of the literature 2007–2017," *Science & Technology Libraries*, vol. 36, no. 3, pp. 235–273, 2017. https://doi.org/10.1080/0194262X.2017.1371658

[3] A. García-Holgado *et al*., "Gender equality in STEM programs: A proposal to analyze the situation of a university about the gender gap," in *2020 IEEE Global Engineering Education Conference (EDUCON)*, 2020, pp. 1824–1830. https://doi.org/10.1109/EDUCON45650.2020.9125326

[4] V. Tandrayen-Ragoobur and D. Gokulsing, "Gender gap in STEM education and career choices: What matters?" *Journal of Applied Research in Higher Education*, vol. 14, no. 3, pp. 1021–1040, 2022. https://doi.org/10.1108/JARHE-09-2019-0235

[5] S. Verdugo-Castro, A. García-Holgado, and M. C. Sánchez-Gómez, "The gender gap in higher STEM studies: A systematic literature review," *Heliyon*, vol. 8, no. 8, p. e10300, 2022. https://doi.org/10.1016/j.heliyon.2022.e10300

[6] M. Vooren, C. Haelermans, W. Groot, and H. M. van den Brink, "Comparing success of female students to their male counterparts in the STEM fields: An empirical analysis from enrollment until graduation using longitudinal register data," *International Journal of STEM Education*, vol. 9, no. 1, pp. 1–17, 2022. https://doi.org/10.1186/s40594-021-00318-8

[7] K. K. Pal, K. Piaget, S. Baller, V. Ratcheva, and S. Zahidi, *Global Gender Gap Report 2022*. World Economic Forum, 2022. https://www.weforum.org/reports/global-gender-gap-report-2022/
[8] N. A. Bowman *et al.*, "Gender representation and academic achievement among STEM-interested students in college STEM courses," *Journal of Research in Science Teaching*, vol. 59, no. 10, pp. 1876–1900, 2022. https://doi.org/10.1002/tea.21778
[9] V. A. Lagesen, I. Pettersen, and L. Berg, "Inclusion of women to ICT engineering–lessons learned," *European Journal of Engineering Education*, vol. 47, no. 3, pp. 467–482, 2022. https://doi.org/10.1080/03043797.2021.1983774
[10] C. Corbett and C. Hill, *Solving the Equation: The Variables for Women's Success in Engineering and Computing*. Washington, DC: American Association of University Women, 2015.
[11] J. Amemiya and L. Bian, "Why are there no girls? Increasing children's recognition of structural causes of the gender gap in STEM," *Cognition*, vol. 245, p. 105740, 2024. https://doi.org/10.1016/j.cognition.2024.105740
[12] S. J. Ceci, D. K. Ginther, S. Kahn, and W. M. Williams, "Women in academic science: A changing landscape," *Psychological Science in the Public Interest*, vol. 15, no. 3, pp. 75–141, 2014. https://doi.org/10.1177/1529100614541236
[13] J. Guo, H. W. Marsh, P. D. Parker, and X. Hu, "Cross-cultural patterns of gender differences in STEM: Gender stratification, gender equality and gender-equality paradoxes," *Educational Psychology Review*, vol. 36, no. 2, p. 37, 2024. https://doi.org/10.1007/s10648-024-09872-3
[14] C. Hill, C. Corbett, and A. St. Rose, *Why so Few? Women in Science, Technology, Engineering, and Mathematics*. Washington, DC: American Association of University Women, 2010.
[15] M. A. Kanny, L. J. Sax, and T. A. Riggers-Piehl, "Investigating forty years of STEM research: How explanations for the gender gap have evolved over time," *Journal of Women and Minorities in Science and Engineering*, vol. 20, no. 2, pp. 127–148, 2014. https://doi.org/10.1615/JWomenMinorScienEng.2014007246
[16] A. Parmaxi *et al.*, "Gender equality in science, technology, engineering and mathematics: Industrial vis-a-vis academic perspective," *Discover Education*, vol. 3, no. 1, p. 3, 2024. https://doi.org/10.1007/s44217-023-00082-7
[17] C. Tomassini, "Gender gaps in science: Systematic review of the main explanations and research agenda," *Education in the Knowledge Society (EKS)*, vol. 22, 2021. https://doi.org/10.14201/eks.25437
[18] UNESCO, *Cracking the Code: Girls' and Women's Education in Science, Technology, Engineering and Mathematics (STEM)*, 2017. https://unesdoc.unesco.org/ark:/48223/pf0000253479
[19] M. T. Wang and J. L. Degol, "Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions," *Educational Psychology Review*, vol. 29, no. 1, pp. 119–140, 2017. https://doi.org/10.1007/s10648-015-9355-x
[20] J. Bossart and N. Bharti, "Women in engineering: Insight into why some engineering departments have more success in recruiting and graduating women," *American Journal of Engineering Education*, vol. 8, no. 2, pp. 127–140, 2017. https://doi.org/10.19030/ajee.v8i2.10070
[21] B. Isaacs, "Mystery of the missing women engineers: A solution," *Journal of Professional Issues in Engineering Education and Practice*, vol. 127, no. 2, pp. 85–91, 2001. https://doi.org/10.1061/(ASCE)1052-3928(2001)127:2(85)
[22] S. J. Correll, "Constraints into preferences: Gender, status, and emerging career aspirations," *American Sociological Review*, vol. 69, no. 1, pp. 93–113, 2004. https://doi.org/10.1177/000312240406900106

[23] J. Eccles, "Gendered educational and occupational choices: Applying the Eccles et al. model of achievement-related choices," *International Journal of Behavioral Development*, vol. 35, no. 3, pp. 195–201, 2011. https://doi.org/10.1177/0165025411398185

[24] E. Jouini, P. Karehnke, and C. Napp, "Stereotypes, underconfidence and decision-making with an application to gender and math," *Journal of Economic Behavior & Organization*, vol. 148, pp. 34–45, 2018. https://doi.org/10.1016/j.jebo.2018.02.002

[25] A. Maries, K. Whitcomb, and C. Singh, "Gender inequities throughout STEM bread-crumb," *Journal of College Science Teaching*, vol. 51, no. 3, pp. 27–36, 2022. https://doi.org/10.1080/0047231X.2022.12290557

[26] C. A. Moss-Racusin, C. Sanzari, N. Caluori, and H. Rabasco, "Gender bias produces gender gaps in STEM engagement," *Sex Roles*, vol. 79, pp. 651–670, 2018. https://doi.org/10.1007/s11199-018-0902-z

[27] T. Schmader, "Gender inclusion and fit in STEM," *Annual Review of Psychology*, vol. 74, pp. 219–243, 2023. https://doi.org/10.1146/annurev-psych-032720-043052

[28] B. Bloodhart, M. M. Balgopal, A. M. A. Casper, L. B. Sample McMeeking, and E. V. Fischer, "Outperforming yet undervalued: Undergraduate women in STEM," *PLoS ONE*, vol. 15, no. 6, p. e0234685, 2020. https://doi.org/10.1371/journal.pone.0234685

[29] B. J. Casad *et al.*, "Gender inequality in academia: Problems and solutions for women faculty in STEM," *Journal of Neuroscience Research*, vol. 99, no. 1, pp. 13–23, 2021. https://doi.org/10.1002/jnr.24631

[30] C. Puritty *et al.*, "Without inclusion, diversity initiatives may not be enough," *Science*, vol. 357, no. 6356, pp. 1101–1102, 2017. https://doi.org/10.1126/science.aai9054

[31] P. Baizán *et al.*, "Activities and technologies to promote women presence in STEAM," *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, vol. 16, no. 3, pp. 241–247, 2021. https://doi.org/10.1109/RITA.2021.3132605

[32] L. J. Bottomley, S. Rajala, and R. Porter, "Women in engineering at North Carolina state university: An effort in recruitment, retention, and encouragement," in *FIE'99 Frontiers in Education. 29th Annual Frontiers in Education Conference. Designing the Future of Science and Engineering Education. Conference Proceedings (IEEE Cat. No. 99CH37011)*, 1999, vol. 1, pp. 11A51–11A53. https://doi.org/10.1109/FIE.1999.839101

[33] T. E. Charlesworth and M. R. Banaji, "Gender in science, technology, engineering, and mathematics: Issues, causes, solutions," *Journal of Neuroscience*, vol. 39, no. 37, pp. 7228–7243, 2019. https://doi.org/10.1523/JNEUROSCI.0475-18.2019

[34] J. García-Peñalvo, A. García-Holgado, A. Dominguez, and J. Pascual, Eds., *Women in STEM in Higher Education: Good Practices of Attraction, Access and Retainment in Higher Education*. Singapore: Springer Singapore, 2022. https://doi.org/10.1007/978-981-19-1552-9

[35] A. Miyake, L. E. Kost-Smith, N. D. Finkelstein, S. J. Pollock, G. L. Cohen, and T. A. Ito, "Reducing the gender achievement gap in college science: A classroom study of values affirmation," *Science*, vol. 330, no. 6008, pp. 1234–1237, 2010. https://doi.org/10.1126/science.1195996

[36] M. M. Jameson, "How mathematical dispositions of adult learners play a role in their STEM choices," *New Directions for Adult and Continuing Education*, vol. 2024, no. 181, pp. 41–49, 2024. https://doi.org/10.1002/ace.20520

[37] N. M. Else-Quest, J. S. Hyde, and M. C. Linn, "Cross-national patterns of gender differences in mathematics: A meta-analysis," *Psychological Bulletin*, vol. 136, no. 1, p. 103, 2010. https://doi.org/10.1037/a0018053

[38] S. Hyde, S. M. Lindberg, M. C. Linn, A. B. Ellis, and C. C. Williams, "Gender similarities characterize math performance," *Science*, vol. 321, no. 5888, pp. 494–495, 2008. https://doi.org/10.1126/science.1160364

[39] S. M. Lindberg, J. Hyde, J. L. Petersen, and M. C. Linn, "New trends in gender and mathematics performance: A meta-analysis," *Psychological Bulletin*, vol. 136, no. 6, pp. 1123–1135, 2010. https://doi.org/10.1037/a0021276

[40] D. Reilly, D. L. Neumann, and G. Andrews, "Investigating gender differences in mathematics and science: Results from the 2011 trends in mathematics and science survey," *Research in Science Education*, vol. 49, no. 1, pp. 25–50, 2019. https://doi.org/10.1007/s11165-017-9630-6

[41] B. M. Casey and C. M. Ganley, "An examination of gender differences in spatial skills and math attitudes in relation to mathematics success: A bio-psycho-social model," *Developmental Review*, vol. 60, p. 100963, 2021. https://doi.org/10.1016/j.dr.2021.100963

[42] S. Meinck and F. Brese, "Trends in gender gaps: Using 20 years of evidence from TIMSS," *Large-Scale Assessments in Education*, vol. 7, no. 1, pp. 1–23, 2019. https://doi.org/10.1186/s40536-019-0076-3

[43] A. M. Steegh, T. N. Höffler, M. M. Keller, and I. Parchmann, "Gender differences in mathematics and science competitions: A systematic review," *Journal of Research in Science Teaching*, vol. 56, no. 10, pp. 1431–1460, 2019. https://doi.org/10.1002/tea.21580

[44] G. Graetz and A. Karimi, "Gender gap variation across assessment types: Explanations and implications," *Economics of Education Review*, vol. 91, p. 102313, 2022. https://doi.org/10.1016/j.econedurev.2022.102313

[45] A. M. Gallagher and J. C. Kaufman, "Gender differences in Mathematics: What we know and what we need to know," in *Gender Differences in Mathematics: An Integrative Psychological Approach*, A. M. Gallagher and J. C. Kaufman, Eds., Cambridge, UK: Cambridge University Press, 2005, pp. 316–331. https://doi.org/10.1017/CBO9780511614446.016

[46] S. Griselda, "Different questions, different gender gap: Can the format of questions explain the gender gap in mathematics?" *mimeo*, 2020.

[47] G. Leder and H. Forgasz, "Measuring who counts: Gender and mathematics assessment," *ZDM Mathematics Education*, vol. 50, no. 4, pp. 687–697, 2018. https://doi.org/10.1007/s11858-018-0939-z

[48] P. O. Saygin, "Gender bias in standardized tests: Evidence from a centralized college admissions system," *Empirical Economics*, vol. 59, pp. 1037–1065, 2020. https://doi.org/10.1007/s00181-019-01662-z

[49] G. Sonnert and M. F. Fox, "Women, men, and academic performance in science and engineering: The gender difference in undergraduate grade point averages," *The Journal of Higher Education*, vol. 83, no. 1, pp. 73–101, 2012. https://doi.org/10.1080/00221546.2012.11777235

[50] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association, 2014.

[51] J. Herman and L. Cook, "Broadening the reach of the fairness standards," in *Fairness in Educational and Psychological Testing: Examining Theoretical, Research, Practice, and Policy Implications of the 2014 Standards*, J. L. Jonson and K. F. Geisinger, Eds., Washington, DC: American Educational Research Association, 2022, pp. 33–59. https://doi.org/10.2307/j.ctv2kzv0fw

[52] K. L. Mulvey, A. J. Hoffman, and L. McGuire, "Fairness and opportunity in stem contexts: Gender, stereotypes, and moral judgments," in *Handbook of Moral Development*, 3rd ed., M. Killen and J. G. Smetana, Eds., New York, NY: Routledge, 2022, pp. 236–252. https://doi.org/10.4324/9781003047247

[53] R. D. Tierney, "Fairness as a multifaceted quality in classroom assessment," *Studies in Educational Evaluation*, vol. 43, pp. 55–69, 2014. https://doi.org/10.1016/j.stueduc.2013.12.003

[54] J. Lee, "Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries," *Learning and Individual Differences*, vol. 19, no. 3, pp. 335–365, 2009. https://doi.org/10.1016/j.lindif.2008.10.009

[55] F. Pajares, "Gender differences in mathematics self-efficacy beliefs," in *Gender Differences in Mathematics: An Integrative Psychological Approach*, A. M. Gallagher and J. C. Kaufman, Eds., Cambridge, UK: Cambridge University Press, 2005, pp. 294–315. https://doi.org/10.1017/CBO9780511614446.015

[56] F. Pajares and M. D. Miller, "Mathematics self-efficacy and mathematical problem solving: Implications of using different forms of assessment," *The Journal of Experimental Education*, vol. 65, no. 3, pp. 213–228, 1997. https://doi.org/10.1080/00220973.1997.9943455

[57] C. Schöber, K. Schütte, O. Köller, N. McElvany, and M. M. Gebauer, "Reciprocal effects between self-efficacy and achievement in mathematics and reading," *Learning and Individual Differences*, vol. 63, pp. 1–11, 2018. https://doi.org/10.1016/j.lindif.2018.01.008

[58] T. Williams and K. Williams, "Self-efficacy and performance in mathematics: Reciprocal determinism in 33 nations," *Journal of Educational Psychology*, vol. 102, no. 2, pp. 453–466, 2010. https://doi.org/10.1037/a0017271

[59] A. L. Zeldin and F. Pajares, "Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers," *American Educational Research Journal*, vol. 37, no. 1, pp. 215–246, 2000. https://doi.org/10.2307/1163477

[60] A. L. Zeldin, S. L. Britner, and F. Pajares, "A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers," *Journal of Research in Science Teaching*, vol. 45, no. 9, pp. 1036–1058, 2008. https://doi.org/10.1002/tea.20195

[61] B. N. Guler, S. Sajadi, A. Alenezi, and D. Bairaktarova, "From doubt to drive: How instructional modality and self-efficacy shape motivation in remedial spatial visualization courses," *International Journal of Engineering Pedagogy*, vol. 14, no. 5, pp. 37–50, 2024. https://doi.org/10.3991/ijep.v14i5.46097

[62] A. Bandura, "Self-efficacy mechanism in human agency," *American Psychologist*, vol. 37, no. 2, pp. 122–147, 1982. https://doi.org/10.1037/0003-066X.37.2.122

[63] S. McNally, "Gender differences in tertiary education: What explains STEM participation?" IZA Policy Paper, no. 165, 2020. https://www.econstor.eu/handle/10419/243451

[64] G. Morán-Soto and O. I. González-Peña, "Mathematics anxiety and self-efficacy of Mexican engineering students: Is there gender gap?" *Education Sciences*, vol. 12, no. 6, p. 391, 2022. https://doi.org/10.3390/educsci12060391

[65] L. Zander, E. Höhne, S. Harms, S. M. Pfost, and M. J. Hornsey, "When grades are high but self-efficacy is low: Unpacking the confidence gap between girls and boys in mathematics," *Frontiers in Psychology*, vol. 11, p. 552355, 2020. https://doi.org/10.3389/fpsyg.2020.552355

[66] L. Wang and Z. Yu, "Gender-moderated effects of academic self-concept on achievement, motivation, performance, and self-efficacy: A systematic review," *Frontiers in Psychology*, vol. 14, p. 1136141, 2023. https://doi.org/10.3389/fpsyg.2023.1136141

[67] C. Huang, "Gender differences in academic self-efficacy: A meta-analysis," *European Journal of Psychology of Education*, vol. 28, no. 1, pp. 1–35, 2013. https://doi.org/10.1007/s10212-011-0097-y

[68] B. Hoffman and G. Schraw, "The influence of self-efficacy and working memory capacity on problem-solving efficiency," *Learning and Individual Differences*, vol. 19, no. 1, pp. 91–100, 2009. https://doi.org/10.1016/j.lindif.2008.08.001

[69] OECD, *PISA 2012 Results: Ready to Learn (Volume III): Students' Engagement, Drive and Self-Beliefs*, PISA. Paris: OECD Publishing, 2013. https://doi.org/10.1787/9789264201170-en

[70] S. J. Spencer, C. M. Steele, and D. M. Quinn, "Stereotype threat and women's math performance," *Journal of Experimental Social Psychology*, vol. 35, no. 1, pp. 4–28, 1999. https://doi.org/10.1006/jesp.1998.1373

[71] A. M. Gallagher, R. De Lisi, P. C. Holst, A. V. McGillicuddy-De Lisi, M. Morely, and C. Cahalan, "Gender differences in advanced mathematical problem solving," *Journal of Experimental Child Psychology*, vol. 75, no. 3, pp. 165–190, 2000. https://doi.org/10.1006/jecp.1999.2532

[72] Z. Zhu, "Gender differences in mathematical problem solving patterns: A review of literature," *International Education Journal*, vol. 8, no. 2, pp. 187–203, 2007. https://hdl.handle.net/2440/44896

[73] A. Gallagher, "Sex differences in problem-solving strategies used by high-scoring examinees on the SAT-M," *ETS Research Report Series*, vol. 1992, no. 1, pp. i–35, 1992. https://doi.org/10.1002/j.2333-8504.1992.tb01464.x

[74] A. Geist and M. King, "Different, not better: Gender differences in mathematics learning and achievement," *Journal of Instructional Psychology*, vol. 35, no. 1, pp. 43–52, 2008. https://psycnet.apa.org/record/2008-04999-007

[75] M. M. Kimball, "A new perspective on women's math achievement," *Psychological Bulletin*, vol. 105, no. 2, pp. 198–214, 1989. https://doi.org/10.1037/0033-2909.105.2.198

[76] R. E. O'Dea, M. Lagisz, M. D. Jennions, and S. Nakagawa, "Gender differences in individual variation in academic grades fail to fit expected patterns for STEM," *Nature Communications*, vol. 9, no. 1, p. 3777, 2018. https://doi.org/10.1038/s41467-018-06292-0

[77] D. Voyer and S. D. Voyer, "Gender differences in scholastic achievement: A meta-analysis," *Psychological Bulletin*, vol. 140, no. 4, pp. 1174–1204, 2014. https://doi.org/10.1037/a0036620

[78] K. Cipora, F. H. Santos, K. Kucian, and A. Dowker, "Mathematics anxiety—where are we and where shall we go?" *Annals of the New York Academy of Sciences*, vol. 1513, no. 1, pp. 10–20, 2022. https://doi.org/10.1111/nyas.14770

[79] R. G. Pizzie and D. J. Kraemer, "The academic anxiety inventory: Evidence for dissociable patterns of anxiety related to math and other sources of academic stress," *Frontiers in Psychology*, vol. 9, p. 2684, 2019. https://doi.org/10.3389/fpsyg.2018.02684

[80] A. Bandura, *Self-Efficacy: The Exercise of Control*. New York, NY, USA: W. H. Freeman and Company, 1997.

[81] D. E. Yupanqui-Lorenzo, E. S. Olivera-Carhuaz, A. A. Reynaga Alponte, V. Pulido-Capurro, C. Carbajal-León, and M. A. Cardoza Sernaqué, "Explanatory model on academic self-efficacy in engineering students: Role of anxiety, dysthymia, and negative affect," *International Journal of Engineering Pedagogy*, vol. 13, no. 5, pp. 91–103, 2023. https://doi.org/10.3991/ijep.v13i5.38577

[82] H. L. Klee, A. D. Miller, and M. M. Buehl, "Mathematics anxiety, self-concept, and self-efficacy: A multidimensional scaling consideration of measures," *The Journal of Experimental Education*, vol. 91, no. 3, pp. 494–516, 2022. https://doi.org/10.1080/00220973.2021.2024788

[83] A. Meador and L. R. Salazar, "The effects of math anxiety on the performance of undergraduate business majors using self-efficacy as a mediator," *International Journal of Teaching and Learning in Higher Education*, vol. 35, no. 2, pp. 147–159, 2023.

[84] D. Szucs and E. Toffalini, "Maths anxiety and subjective perception of control, value and success expectancy in mathematics," *Royal Society Open Science*, vol. 10, no. 11, p. 231000, 2023. https://doi.org/10.1098/rsos.231000

[85] N. Hendral and K. Hidayati, "The relationship between students' self-efficacy and mathematics anxiety: Meta-analysis investigation," *AIP Conference Proceedings*, vol. 2540, no. 1, 2023. https://doi.org/10.1063/5.0105860

[86] N. E. Betz, "Prevalence, distribution, and correlates of math anxiety in college students," *Journal of Counseling Psychology*, vol. 25, no. 5, pp. 441–448, 1978. https://doi.org/10.1037/0022-0167.25.5.441

[87] S. Caviola, E. Toffalini, D. Giofrè, J. M. Ruiz, D. Szűcs, and I. C. Mammarella, "Math performance and academic anxiety forms, from sociodemographic to cognitive aspects: A meta-analysis on 906,311 participants," *Educational Psychology Review*, vol. 34, pp. 363–399, 2022. https://doi.org/10.1007/s10648-021-09618-5

[88] A. Devine, K. Fawcett, D. Szűcs, and A. Dowker, "Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety," *Behavioral and Brain Functions*, vol. 8, p. 33, 2012. https://doi.org/10.1186/1744-9081-8-33

[89] R. Hembree, "The nature, effects, and relief of mathematics anxiety," *Journal for Research in Mathematics Education*, vol. 21, no. 1, pp. 33–46, 1990. https://doi.org/10.3102/00346543058001047

[90] M. J. Justicia-Galiano, M. E. Martín-Puga, R. Linares, and S. Pelegrina, "Gender stereotypes about math anxiety: Ability and emotional components," *Learning and Individual Differences*, vol. 105, p. 102316, 2023. https://doi.org/10.1016/j.lindif.2023.102316

[91] S. Rossi *et al.*, "Mathematics–gender stereotype endorsement influences mathematics anxiety, self-concept, and performance differently in men and women," *Annals of the New York Academy of Sciences*, vol. 1513, no. 1, pp. 121–139, 2022. https://doi.org/10.1111/nyas.14779

[92] C. Barroso, C. M. Ganley, A. L. McGraw, E. A. Geer, S. A. Hart, and M. C. Daucourt, "A meta-analysis of the relation between math anxiety and math achievement," *Psychological Bulletin*, vol. 147, no. 2, pp. 134–168, 2021. https://doi.org/10.1037/bul0000307

[93] X. Ma, "A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics," *Journal for Research in Mathematics Education*, vol. 30, no. 5, pp. 520–540, 1999. https://doi.org/10.2307/749772

[94] J. M. Namkung, P. Peng, and X. Lin, "The relation between mathematics anxiety and mathematics performance among school-aged students: A meta-analysis," *Review of Educational Research*, vol. 89, no. 3, pp. 459–496, 2019. https://doi.org/10.3102/0034654319843494

[95] Zhang, N. Zhao, and Q. P. Kong, "The relationship between math anxiety and math performance: A meta-analytic investigation," *Frontiers in Psychology*, vol. 10, p. 1613, 2019. https://doi.org/10.3389/fpsyg.2019.01613

[96] M. H. Ashcraft and E. P. Kirk, "The relationships among working memory, math anxiety, and performance," *Journal of Experimental Psychology: General*, vol. 130, no. 2, pp. 224–237, 2001. https://doi.org/10.1037/0096-3445.130.2.224

[97] D. R. Hopko, D. W. McNeil, C. W. Lejuez, M. H. Ashcraft, G. H. Eifert, and J. Riel, "The effects of anxious responding on mental arithmetic and lexical decision task performance," *Journal of Anxiety Disorders*, vol. 17, no. 6, pp. 647–665, 2003. https://doi.org/10.1016/S0887-6185(02)00240-2

[98] M. H. Ashcraft and J. A. Krause, "Working memory, math performance, and math anxiety," *Psychonomic Bulletin & Review*, vol. 14, pp. 243–248, 2007. https://doi.org/10.3758/BF03194059

[99] M. H. Ashcraft and A. M. Moore, "Mathematics anxiety and the affective drop in performance," *Journal of Psychoeducational Assessment*, vol. 27, no. 3, pp. 197–205, 2009. https://doi.org/10.1177/0734282908330580

[100] J. Finell. E. Sammallahti, J. Korhonen, H. Eklöf, and B. Jonsson, "Working memory and its mediating role on the relationship of math anxiety and math performance: A meta-analysis," *Frontiers in Psychology*, vol. 12, p. 798090, 2022. https://doi.org/10.3389/fpsyg.2021.798090

[101] D. R. Hopko, M. H. Ashcraft, J. Gute, K. J. Ruggiero, and C. Lewis, "Mathematics anxiety and working memory: Support for the existence of a deficient inhibition mechanism," *Journal of Anxiety Disorders*, vol. 12, no. 4, pp. 343–355, 1998. https://doi.org/10.1016/S0887-6185(98)00019-X

[102] A. Arenas and C. Calsamiglia, "Gender differences in high-stakes performance and college admission policies," Discussion Paper Series No. 15550, Institute of Labor Economics, 2022. https://docs.iza.org/dp15550.pdf

[103] G. Azmat, C. Calsamiglia, and N. Iriberri, "Gender differences in response to big stakes," *Journal of the European Economic Association*, vol. 14, no. 6, pp. 1372–1400, 2016. https://doi.org/10.1111/jeea.12180

[104] B. Borges, F. Estevan, and L. P. Morin, "Gender differences in prioritizing rewarding tasks," *SSRN*, 2023. https://doi.org/10.2139/ssrn.4354209

[105] X. Cai, Y. Lu, J. Pan, and S. Zhong, "Gender gap under pressure: Evidence from China's national college entrance examination," *Review of Economics and Statistics*, vol. 101, no. 2, pp. 249–263, 2019. https://doi.org/10.1162/rest_a_00749

[106] E. Lyons, A. Mesghina, and L. E. Richland, "Complicated gender gaps in mathematics achievement: Elevated stakes during performance as one explanation," *Mind, Brain, and Education*, vol. 16, no. 1, pp. 36–47, 2022. https://doi.org/10.1111/mbe.12312

[107] D. Montolio and P. A. Taberner, "Gender differences under test pressure and their impact on academic performance: A quasi-experimental design," *Journal of Economic Behavior & Organization*, vol. 191, pp. 1065–1090, 2021. https://doi.org/10.2139/ssrn.3296211

[108] A. Schlosser, Z. Neeman, and Y. Attali, "Differential performance in high versus low stakes tests: Evidence from the GRE test," *The Economic Journal*, vol. 129, no. 623, pp. 2916–2948, 2019. https://doi.org/10.1093/ej/uez015

[109] L. Radević and I. Milovanović, "Current trends in math anxiety research: A bibliometric approach," *International Journal of Science and Mathematics Education*, vol. 22, no. 6, pp. 1345–1362, 2024. https://doi.org/10.1007/s10763-023-10424-4

[110] R. D. Tierney, "Fairness in classroom assessment," in *Handbook of Research on Classroom Assessment*, J. H. McMillan, Ed., Thousand Oaks, CA: SAGE, 2013, pp. 125–144. https://doi.org/10.4135/9781452218649

[111] A. Rasooli and C. DeLuca, "A critical review of fairness from multiple perspectives: Implications for classroom assessment theory," *Applied Measurement in Education*, vol. 37, no. 2, pp. 148–164, 2024. https://doi.org/10.1080/08957347.2024.2345594

[112] M. Azizi, M. Imad, S. M. Foote, J. Caulkins, and B. Wuetherick, "Humanizing STEM education: An exploratory study of faculty approaches to course redesign," *Frontiers in Education*, vol. 8, p. 1181157, 2023. https://doi.org/10.3389/feduc.2023.1181157

[113] C. T. Amelink and P. S. Meszaros, "A comparison of educational factors promoting or discouraging the intent to remain in engineering by gender," *European Journal of Engineering Education*, vol. 36, no. 1, pp. 47–62, 2011. https://doi.org/10.1080/03043797.2010.531695

[114] D. Khodadad, "Creating a supportive and effective learning environment for engineering students: Pedagogical strategies, engagement, and enhanced outcomes," *International Journal of Engineering Pedagogy*, vol. 13, no. 8, pp. 33–50, 2023. https://doi.org/10.3991/ijep.v13i8.41755

[115] T. T. A. Ngo, "Perception of engineering students on social constructivist learning approach in classroom," *International Journal of Engineering Pedagogy*, vol. 14, no. 1, pp. 20–38, 2024. https://doi.org/10.3991/ijep.v14i1.43101

[116] E. J. Theobald *et al.*, "Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math," *Proceedings of the National Academy of Sciences*, vol. 117, no. 12, pp. 6476–6483, 2020. https://doi.org/10.1073/pnas.1916903117

[117] S. Beltozar-Clemente, O. Iparraguirre-Villanueva, J. Zapata-Paulini, and M. Cabanillas-Carbonell, "Changing mathematical paradigms at the university level: Feedback from a flipped classroom at a Peruvian university," *International Journal of Engineering Pedagogy*, vol. 13, no. 6, pp. 76–89, 2023. https://doi.org/10.3991/ijep.v13i6.40763

[118] A. González-Cervera and Y. González Arechavala, "Estudios STEM en España y participación de la mujer. La Formación Profesional STEM, una oportunidad de futuro," Cátedra para la Promoción de la Mujer en vocaciones STEM en la Formación Profesional para la Movilidad Sostenible, 2021. https://files.griddo.comillas.edu/estudios-stem-en-espac3b1a-y-participacion-de-la-mujer-dic-21.pdf

[119] M. Schneider and F. Preckel, "Variables associated with achievement in higher education: A systematic review of meta-analyses," *Psychological Bulletin*, vol. 143, no. 6, pp. 565–600, 2017. https://doi.org/10.1037/bul0000098

[120] G. M. Garbanzo, "Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública," *Revista Educación*, vol. 31, no. 1, pp. 43–63, 2007. https://doi.org/10.15517/revedu.v31i1.1252

[121] C. Martínez de Ibarreta, A. Rúa, R. Redondo, M. E. Fabra, A. Núñez, and M. J. Martín, "Influencia del nivel educativo de los padres en el rendimiento académico de los estudiantes de la ADE. Un enfoque de género," in *Investigaciones de Economía de la Educación Número 5*, AEDE, Asociación de Economía de la Educación, 2010, pp. 1273–1296. https://2010.economicsofeducation.com/user/pdfsesiones/113.pdf

[122] J. L. Arroyo-Barrigüete, G. Tirado, I. Mahíllo-Fernández, and P. J. Ramírez, "Predictores del rendimiento académico en las titulaciones de Administración y Dirección de Empresas: el efecto de la especialidad en bachillerato," *Revista de Educación*, vol. 390, pp. 129–154, 2020. https://doi.org/10.4438/1988-592X-RE-2020-390-468

[123] K. McKenzie and R. Schweitzer, "Who succeeds at university? Factors predicting academic performance in first year Australian university students," *Higher Education Research & Development*, vol. 20, no. 1, pp. 21–33, 2001. https://doi.org/10.1080/07924360120043621

[124] A. Beltrán and K. La Serna, "¿Qué explica el rendimiento académico en el primer año de estudios universitarios? Un estudio de caso en la Universidad del Pacífico," Working Paper No. DD/08/09, 2008. https://repositorio.up.edu.pe/item/b2bcfb3b-0b95-4610-87b0-4d1942c970e4

[125] F. J. Tejedor, "Poder explicativo de algunos determinantes del rendimiento en los estudios universitarios," *Revista Española de Pedagogía*, vol. 224, pp. 5–32, 2003.

[126] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2025. https://www.R-project.org/

[127] A. Zeileis and T. Hothorn, "Diagnostic checking in regression relationships," *R News*, vol. 2, no. 3, pp. 7–10, 2002. https://CRAN.R-project.org/doc/Rnews/

[128] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Cham: Springer-Verlag, 2016. https://doi.org/10.1007/978-3-319-24277-4_9

[129] M. Nakazawa, *fmsb: Functions for Medical Statistics Book with some Demographic Data*, R Package Version 0.7.0, 2019. https://CRAN.R-project.org/package=fmsb

[130] T. Lang and CRAN team, *Package 'RCurl'*, R Package Version 1.98-1.2, 2019. https://CRAN.R-project.org/package=RCurl

[131] J. Fox and S. Weisberg, *An R Companion to Applied Regression, 3rd, Ed.,* Thousand Oaks, CA: Sage, 2018.

[132] P. Mair and R. R. Wilcox, "Robust statistical methods in R using the WRS2 package," *Behavior Research Methods*, vol. 52, pp. 464–488, 2020. https://doi.org/10.3758/s13428-019-01246-w

[133] M. Kuhn, *Caret: Classification and Regression Training*, R Package Version 6.0-86, 2020. https://CRAN.R-project.org/package=caret

[134] H. Wickham, R. François, L. Henry, and K. Müller, *dplyr: A Grammar of Data Manipulation*, R Package Version 1.0.4, 2021. https://CRAN.R-project.org/package=dplyr
[135] S. Gaure, *lfe: Linear Group Fixed Effects*, R Package Version 2.8-6, 2021. https://CRAN.R-project.org/package=lfe
[136] A. Kassambara, *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*, R Package Version 0.7.2, 2023. https://CRAN.R-project.org/package=rstatix
[137] J. Pizarroso, J. Portela, and A. Muñoz, "NeuralSens: Sensitivity analysis of neural networks," *Journal of Statistical Software*, vol. 102, no. 7, pp. 1–36, 2022. https://doi.org/10.18637/jss.v102.i07
[138] H. J. R. Terano, "Exploring the academic preparation and performance to the comprehensive examination of first-year engineering students," *International Journal of Engineering Pedagogy*, vol. 13, no. 7, pp. 116–127, 2023. https://doi.org/10.3991/ijep.v13i7.41217
[139] J. S. Brotman and F. M. Moore, "Girls and science: A review of four themes in the science education literature," *Journal of Research in Science Teaching*, vol. 45, no. 9, pp. 971–1002, 2008. https://doi.org/10.1002/tea.20241
[140] A. M. Steegh, T. Höffler, L. Höft, and I. Parchmann, "First steps toward gender equity in the chemistry Olympiad: Understanding the role of implicit gender-science stereotypes," *Journal of Research in Science Teaching*, vol. 58, no. 1, pp. 40–68, 2021. https://doi.org/10.1002/tea.21645
[141] D. Rozgonjuk, T. Kraav, K. Mikkor, K. Orav-Puurand, and K. Täht, "Mathematics anxiety among STEM and social sciences students: The roles of mathematics self-efficacy, and deep and surface approach to learning," *International Journal of STEM Education*, vol. 7, no. 1, pp. 1–11, 2020. https://doi.org/10.1186/s40594-020-00246-z
[142] B. Bordel, R. Alcarria, T. Robles, and D. Martin, "The gender gap in engineering education during the COVID-19 lockdown: A study case," *International Journal of Engineering Pedagogy*, vol. 11, no. 6, pp. 117–131, 2021. https://doi.org/10.3991/ijep.v11i6.24945

9 APPENDIX

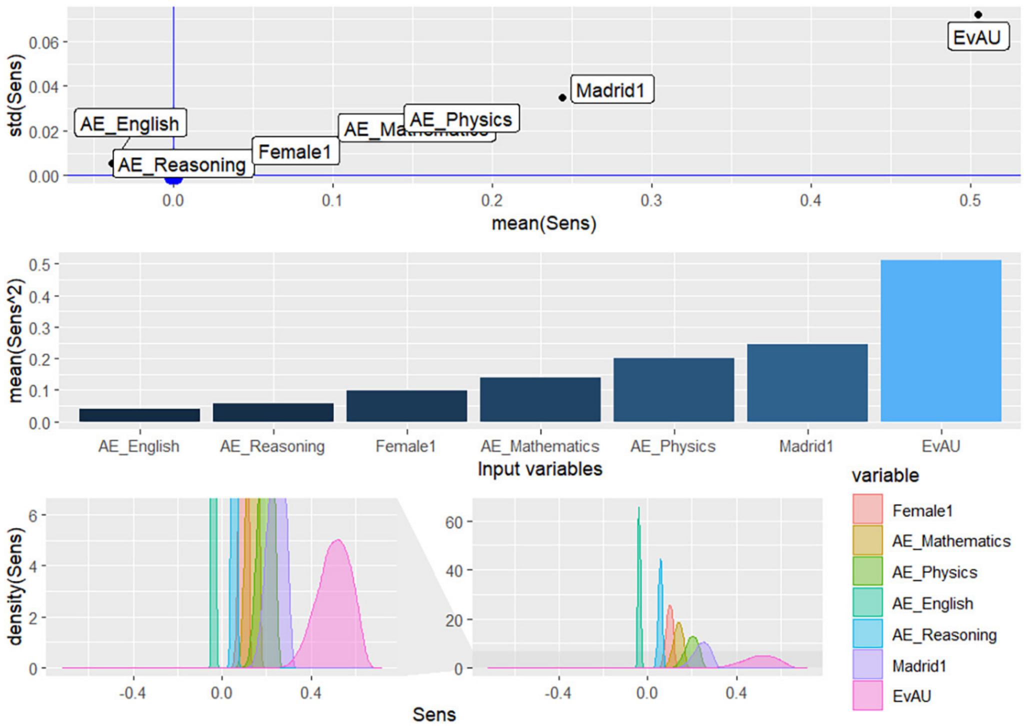

Fig. A1. NeuralSens graphical results (Algebra)

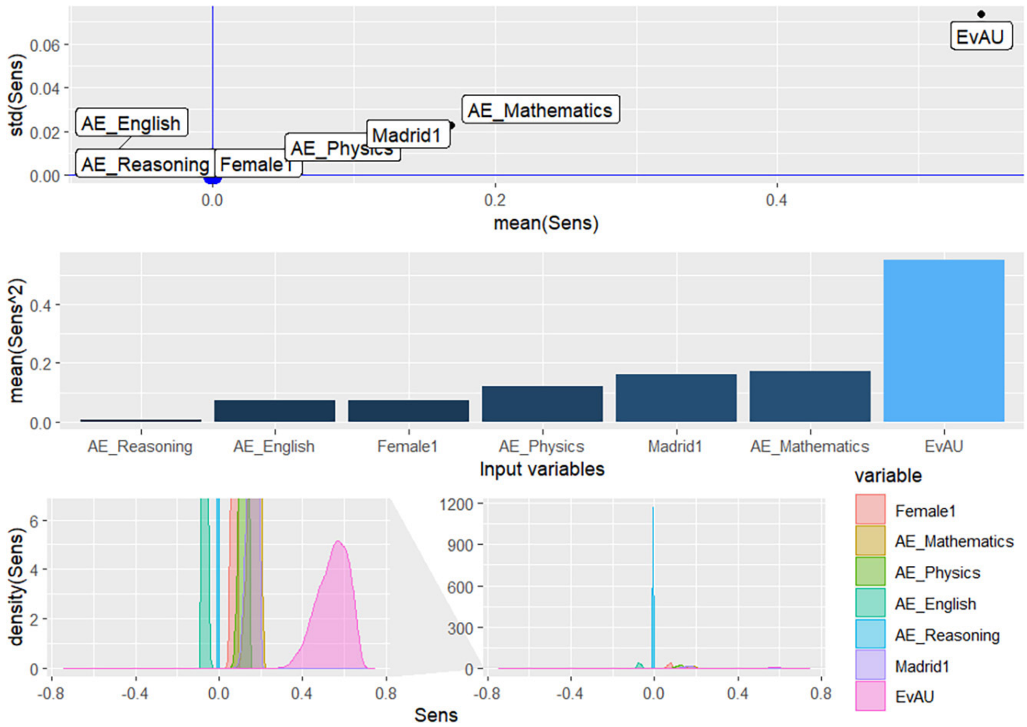

Fig. A2. NeuralSens graphical results (Calculus)

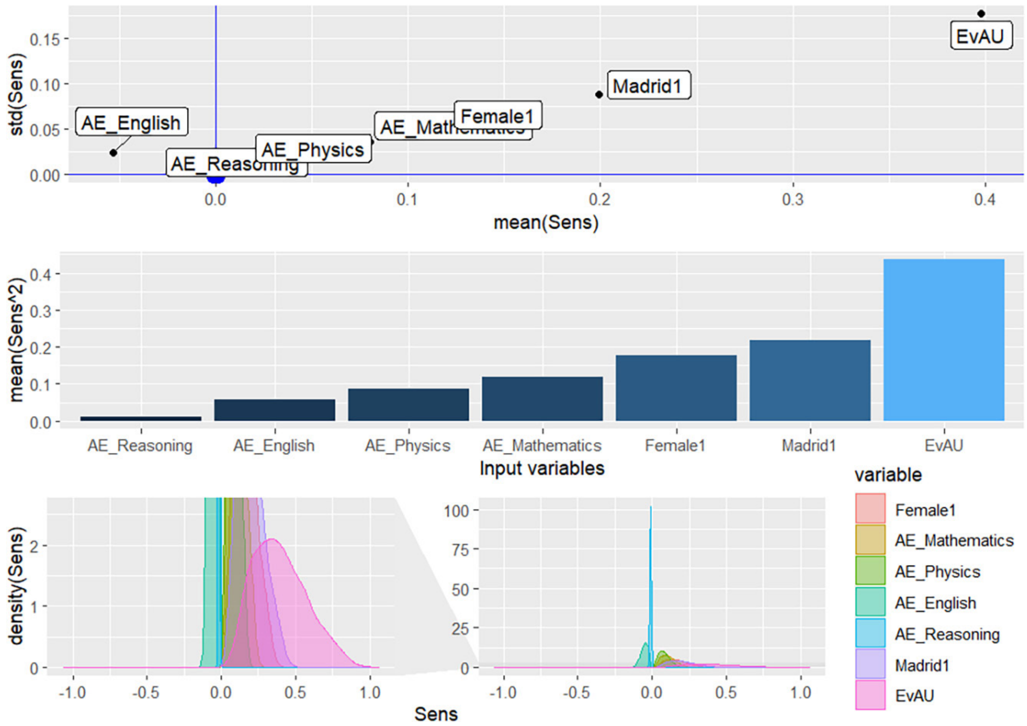

Fig. A3. NeuralSens graphical results (Differential Equations)

10 AUTHORS

Jose Luis Arroyo-Barrigüete is an Associate Professor at the Business School, Universidad Pontificia Comillas, and Co-Director of the Santalucía Chair of Analytics for Education. He holds a PhD in Business Administration and a PhD in Engineering. His main area of research is teaching innovation (E-mail: jlarroyo@comillas.edu).

Susana Carabias-López is an Associate Professor at the Business School, Universidad Pontificia Comillas. She holds a PhD in Business Administration, and her main areas of research are teaching innovation and corporate finance.

Antonio Obregón is a Full Professor in the Department of Public Law at Pontifical University of Comillas. He holds a PhD in Criminal Law and a Degree in Business Administration. He is a Former Vice Rector of Academic Affairs and Faculty, Director of Madrid Culinary Campus, and a Member of the Santalucía Chair of Analytics for Education.

Yolanda González-Arechavala, PhD, is an Associate Professor of the Department of Telematics and Computer Science and the Institute for Research in Technology (IIT), Universidad Pontificia Comillas, Madrid, Spain and the Director of the Chair for the Promotion of Women in STEM in Vocational Training for Sustainable Mobility – Universidad Pontificia Comillas.

Santiago Canales-Cano studied at the Complutense University of Madrid, where he graduated with a degree in Mathematical Sciences in 1994 and obtained his PhD in Computer Science from the Polytechnic University of Madrid in 2004. Since October 1996, he has been a Professor in the Department of Applied Mathematics at the ETSI-ICAI of the Universidad Pontificia Comillas. Since 1997 he has been developing his research activity in the field of Computational Geometry and Discrete Mathematics and more specifically, in heuristic techniques for geometric problems of NP-hard nature.