

A Gender Gap Analysis on Academic Performance in Engineering Students on Admission and Exit Standardized Tests

Análisis de brechas de género en el desempeño académico en pruebas de admisión y egreso en programas de Ingeniería

Luis E. Gallego¹, María A. Casadiego²

ABSTRACT

This paper describes a gender-based research work on academic performance in engineering students. This study is based on the results of 9 469 students from Universidad Nacional de Colombia in the college's Admission and exit standard tests (the latter known as Saber Pro Tests). Tools such as descriptive statistics, regression analysis, and data mining are used to estimate both gaps and leaps in scores per gender. These tools are not only used to estimate said gaps, but also to determine whether these gaps are broadened or closed throughout the university education process. The results show that there are still gender gaps in favor of men in the area of mathematics, even in STEM programs. On the contrary, a gap in favor of women is noted in the area of writing, although a decline is also observed in reading comprehension skills for both genders. On the other hand, in terms of the global scores, women improve more than men as a result of their undergraduate experience. Finally, purely disciplinary competencies exhibit a notorious gender gap in favor of men, which should lead to future reforms in this type of programs.

Keywords: gender gap, academic performance, standardized tests, regression analysis, data mining

RESUMEN

Este artículo describe una investigación sobre brechas de género en el desempeño académico de los estudiantes en programas de ingeniería. El estudio está basado en los resultados de 9 469 estudiantes de la Universidad Nacional de Colombia en las pruebas de admisión y egreso (estas últimas conocidas como pruebas Saber Pro). Se utilizan herramientas como la estadística descriptiva, el análisis de regresión y la minería de datos para estimar tanto brechas como saltos de puntaje por género, las cuales se utilizan no solo para estimar tales brechas, sino para determinar si las mismas han sido ampliadas o cerradas a lo largo del proceso de formación universitaria. Los resultados muestran que aún persisten brechas de género en favor de los hombres en el área de matemáticas, incluso en programas STEM. Por el contrario, es notoria una brecha en favor de las mujeres en el área de escritura, aunque también se observa un deterioro en las habilidades de comprensión de lectura para ambos géneros. Por otra parte, en términos de puntajes globales, las mujeres mejoran más que los hombres como resultado de la experiencia del pregrado. Finalmente, las competencias puramente disciplinares exhiben una notoria brecha de género en favor de los hombres, lo que debe llevar a futuras reformas en este tipo de cursos.

Palabras clave: brecha de género, desempeño académico, pruebas estandarizadas, análisis de regresión, minería de datos

Received: June 21st 2022

Accepted: May 18th 2023

Introduction

According to the *2020 Global Gender Gap Report*, gaps in educational achievement still require a decisive effort from worldwide governments to close them. Only 35 countries have achieved gender parity in education and 10% of girls from 15 to 24 years old around world are still illiterate, most of them concentrated in developing countries (WEF, 2020). Consequently, more investment in human capital is still required to improve poor academic performance. In addition, even in countries with high educational achievement, women's abilities are not necessarily aligned with the required skills to make them successful in the so-called *future professions* (WEF, 2020).

Gender gaps are pronounced in college STEM programs, as only 35% of the enrolled students are women, and there is a decreasing trend in their participation in areas such as engineering, natural sciences, mathematics, statistics, and ICTs* (UNESCO, 2019).

As mentioned by Alam (2022), there are three main approaches that explain gender gaps in STEM education: biological, psychological, and sociocultural. On the one hand, the biological approach is based on cognitive patterns, brain functioning (Kucian *et al.*, 2005), and visual-spatial skills (Harris *et al.*, 2021; Reilly *et al.*, 2017; Lawton and Fletcher, 2005). However, much of the recent research argues that gender inequalities in STEM career choices are best explained by relative cognitive strengths rather than cognitive ability levels alone (Tandrayen and

¹ Electrical Engineer, Universidad Nacional de Colombia, Colombia. M.Sc Electrical Engineering, Universidad Nacional de Colombia, Colombia. PhD in Electrical Engineering, Universidad Nacional de Colombia, Colombia. Affiliation: Full professor at the Electrical and Electronics Engineering Department, School of Engineering, Bogotá Campus Universidad Nacional de Colombia. Email: lgallegov@unal.edu.co.

² BSc in Literary Studies, Universidad Nacional de Colombia, Colombia. MSc in Literary Studies, Universidad Nacional de Colombia, Colombia. Affiliation: Advisor at the Vice-Ministry of Elementary and Secondary Education, Ministry of Education of Colombia. Email: macasadiegos@unal.edu.co. ORCID:



Attribution 4.0 International (CC BY 4.0) Share - Adapt

*Information and Communication Technologies

Gokulsing, 2021). On the other hand, some authors argue that psychological and sociocultural contexts determine the expectancy of women's future participation in STEM education due to (i) gender stereotypes (García-Holgado *et al.*, 2020; Makarova *et al.*, 2019; UNESCO, 2019; García-Holgado *et al.*, 2019; Carlana, 2019; Verdugo-Castro *et al.*, 2018; Riegle-Crumb, 2005), (ii) beliefs regarding abilities (Kuschel *et al.*, 2020; Jungert *et al.*, 2019), and (iii) social interactions (Brenøe and Zolitz, 2020; Raabe *et al.*, 2019).

Colombia has implemented several initiatives to close both gaps, *i.e.*, in learning and egalitarian access to education. Since 2012, public education is free from Elementary to High School. However, (i) socioeconomic background, (ii) regional and ethnic provenance, and (iii) gender are factors that still restrict access to education (MEN and OECD, 2016). In addition, Colombia is not far from the world gender-gap trend in STEM programs. Between 2001 and 2018, more often than men, women chose curricular programs historically associated with female roles, such as health and education sciences. Moreover, even though the gender gap is smaller, this trend was also observed in social and human sciences. In contrast, men are more prone to choosing programs such as engineering, architecture, or urban sciences (CPEM *et al.*, 2020). Consequently, many government and education entities have undertaken some institutional efforts to close this gender gap and improve diversity in engineering education, which may demand a change in some education approaches in the context of the future challenges stated in the Sustainable Development Goals (UNESCO, 2019).

Regarding academic performance, some standardized tests have shown a notorious gender gap in some areas such as science and mathematics in Colombia. For instance, the PISA 2018 math test results showed that the gender gap is 20 points in favor of men, while, in the OECD countries, this value is only 5 points. In contrast, the gender gap in science in the OECD countries is 2 points in favor of women, but, in Colombia, this gap is 12 points in favor of men (Radinger *et al.*, 2018). In addition, Abadía and Bernal (2017) found that women underperform men with similar characteristics in the Colombian High School Exit Examination (known as the *Saber 11 Test*), where personal, family, and school characteristics explain a small part of the gap. However, differences in returns and regional provenance are very determinant. A similar regional approach in the *Saber 11* test is found in Cárcamo and Mola (2019). On the other hand, Rincón and Arias (2019) found that, in the College Exit Examination (known as the *Saber Pro Test*), men from Administration areas in Higher Education programs not only score higher in quantitative reasoning but also in English (second language), reading and citizenship competencies, which are language-related areas where women usually score higher than men.

Based on the literature review, gender gap approaches are mainly focused on the Elementary and High School levels, where mathematics and sciences are the most critical areas, as they show increasing gender gaps. This paper hopes to contribute to analyzing gender gaps, this time not only in Higher Education, but also in STEM programs, where closing such gaps is expected, given that the students admitted to these programs are supposedly skillful in math and science. In addition, this study analyzes the change or leap in the

academic performance of men and women between the entrance (admission) and exit exams, so that (i) the effect of university education may be estimated and (ii) the gender gap's behavior may be assessed.

This paper is organized as follows. First, a description of the data is presented in terms of the target population and competencies. Next, some methodological approaches are described in terms of the proposed variables and the methods of analysis. Then, the main results are presented in terms of descriptive statistics, regression analysis, and the data mining approach adopted. Finally, the main conclusions of the study are presented.

Data

This study was conducted over a five-year period (2016-2020), based on the scores obtained by 9469 engineering students on the *Saber Pro* test. The data were collected from two main sources: (i) official records from the ICFES testing agency, and (ii) official admission results from Universidad Nacional de Colombia's Admissions Office. Academic information regarding program enrollment was obtained from ICFES, and socioeconomic information such as strata and public or private schooling was obtained from the official records of the Admissions Office. In addition, students' identities were collected from ICFES, such as biological sex. It is worth noting that the term *gender* refers to biological sex, as gender identity data were not collected.

Population

The students analyzed came from four different campuses and 31 different engineering programs of Universidad Nacional de Colombia, *i.e.*, Bogotá (9), Medellín (13), Manizales (6) and Palmira (3), covering a wide range of disciplinary areas such as Civil, Chemical, Electrical and Electronics, Mechanical and Mechatronics, Industrial and Computer, Geological, Mining, Agricultural, Environmental, and Petroleum Engineering, among others. A previous study was carried out by Gallego (2021) for the entire population of the same university, using only descriptive statistics. This research is now gender-oriented and exclusively focused on engineering programs. In addition, not only does it use descriptive statistics, but also inferential statistics and data mining approaches.

The gender composition is shown in Figure 1, where an approximate 75/25% is the dominant ratio between men and women on all campuses but Palmira, which shows a more balanced ratio (51/49%). In addition, Figures 2 and 3 show that social strata are concentrated in the low-middle class (strata 2 and 3, with about 75%), as well as a well-balanced ratio between public and private schooling (52/48%). Both the social stratum and the schooling type show no significant gender gaps.

Competencies

In general, the *Saber Pro* Test evaluates two types of competencies: general and disciplinary. General competencies refer to the skills that must be developed by students regardless of their background field (ICFES, 2021), which are evaluated according to the following modules:

[†]OECD: Organisation for Economic Co-operation and Development

- Quantitative reasoning (Quant)
- English as a Second language test (Eng)
- Written communication (Wrt)
- Reading comprehension (Read)
- Citizenship competencies (Cty)

On the other hand, disciplinary competencies are chosen by each curricular program from a set proposed for each knowledge area. In particular, the knowledge area for this study was engineering, which includes 15 specific disciplinary skills (Figure 8).

Methodology

Variables

Gender Gap Scores (GGS)

In general, the main purpose of this study is to analyze the differences between women and men’s scores. These differences are dubbed *Gender Gap Scores* (GGS) and are treated via two different statistical approaches: parametric and non-parametric.

Regarding the parametric approach, GGS are defined as the difference in means between women and men’s scores, as shown in Equation (1). A positive/negative GGS value means that women/men scored higher than men/women in the Saber Pro Test. However, even though both means are different, some tests must be run in order to determine whether these differences are statistically significant. This study used a T-test to compare both population means while fulfilling some conditions such as (i) independence, (ii) normality and (iii) homoscedasticity. In this case, the observations were assumed to be independent, given that the exam is taken individually by each student. In addition, normality and homoscedasticity assumptions were respectively verified via Shapiro-Wilks and Fligner-Killeen tests, using the Welch correction in the T-test for the cases exhibiting heteroscedasticity. Finally, a D-Cohen index was calculated to estimate the effect size of the GGS. These estimates are presented in Table 1.

$$GGS = \overline{Score_{Women}} - \overline{Score_{Men}} \quad (1)$$

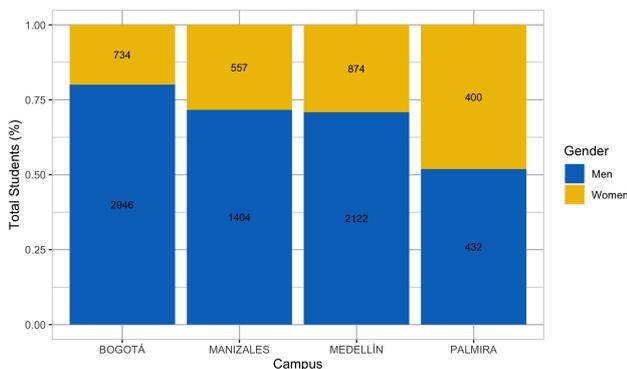


Figure 1. Percent gender distribution and number of students per campus
Source: Authors

Now, to validate the parametric approach, if the T-test applied in samples larger than 30 still showed them to be normally distributed, a bootstrapping method was used to estimate the GGS. Therefore, a resampling with 500 stratified bootstrapped samples (women-men) and a 95% confidence interval was used to estimate the mean of the GGS. These estimates are shown in Table 2. In general, both approaches, parametric and non-parametric, led to almost identical GGS values.

Gender Leap Scores (GLS)

The main purpose of Gender Leap Scores (GLS) is to analyze the changes in academic performance between an entrance and an exit exam, so that the effect of university education can be estimated. In this case, academic performance may increase or decrease as a consequence of obtaining positive or negative GLS, respectively. In this study, the entrance exam corresponds to the University’s admission test, while the exit exam is the Saber Pro Test, often taken during the final year of the curriculum.

The scores obtained in both exams must be normalized in order for them to be comparable. Therefore, a max-min normalization was performed while following Equation (2).

$$Score_{esc} = 100 * \frac{(Score - min(score))}{(max(score) - min(score))} \quad (2)$$

where:

- $Score_{esc}$ corresponds to an escalated score that *only* refers to the students that took the Saber Pro Test.
- $Score$ corresponds to the score obtained in the Saber Pro or the admission test.
- $min(score)/max(score)$ corresponds to the minimum/maximum Saber Pro or admission test score only obtained by the students that took the Saber Pro test.

It is important to note that $Score_{esc} \in [0, 100]$ is only relative to the students that took both tests, so a comparison between both (admission and Saber Pro escalated scores) may describe a change in the academic performance by gender (*i.e.*, a GLS).

‡Statistical indicator comparing the difference in means against the joint standard deviation of both samples

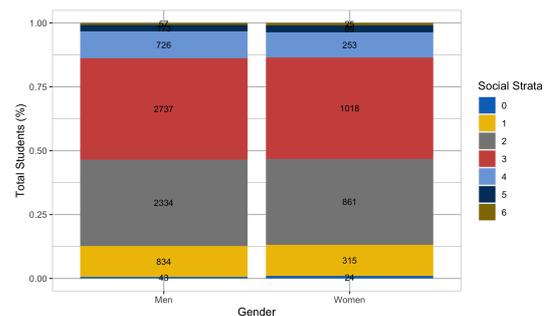


Figure 2. Percent social stratum distribution and number of students per gender
Source: Authors

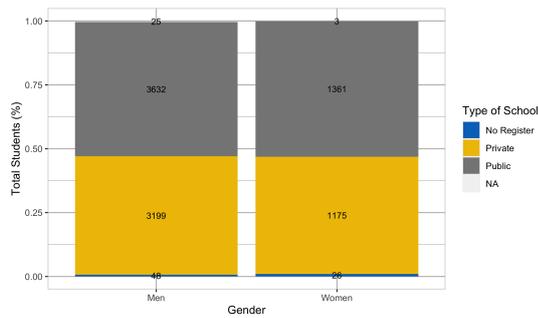


Figure 3. Percent schooling type distribution and number of students per gender
Source: Authors

On the other hand, a plausible comparison between the entrance and exit tests must consider only commonly assessed competencies or skills. In particular, the admission test assesses only the general skills acquired in high-school, such as quantitative reasoning and reading comprehension, but also natural sciences, social sciences, and image analysis. In contrast, the exit exam assesses both general (quantitative reasoning, reading comprehension, written communication, citizenship, and English) and disciplinary competencies (scientific thinking, design, and project formulation, as shown in Figure 8). Therefore, it is not possible to assess a leap in areas that were not specifically assessed in both exams, such as citizenship, English, or written communication. Thus, GLS were calculated only for quantitative reasoning and reading comprehension. In addition, a GLS was calculated for the global score, as a gender gap in global scores was observed in the admission test.

Types of analysis

The types of analysis performed on GGS and GLS were the following:

1. Descriptive statistics analysis: behavior of GGS and GLS score means, confidence intervals, interquartile ranges, and statistical significance tests per skill.
2. Regression analysis
 - (a) Correlograms: correlation coefficients between different skill scores in both tests per gender.
 - (b) Linear multivariate analysis: linear regressions on global scores explained by the different skill scores per gender. The regressions performed follow Equation (4).
 - (c) Logistic regression analysis: logistic regressions on high/low scores or positive/negative leaps regarding gender via Equation (5). The main purpose is to determine the Odds ratios between men and women on the different skill scores.

3. Data mining analysis (association rules): the main purpose of association rules is to discover hidden relations between scores and other dimensions (gender, social strata, and schooling type) based on their frequency of appearance. In this study, a basic Apriori algorithm was applied to obtain simple if-then rules in the form $X \Rightarrow Y$. Only strong rules were analyzed based on support and confidence indicators. Generally speaking, support denotes the frequency of appearance of an itemset X in relation to the total of transactions. In addition, confidence refers to the number of times that if-then rules appear, following Equation (3).

$$Confidence(X \Rightarrow Y) = \frac{Support(X \cup Y)}{Support(X)} \quad (3)$$

where $X \cup Y$ is an itemset including all items in X and Y . Confidence may be interpreted as a conditional probability $P(Y|X)$, i.e, the probability of an item including X while also including Y . In this study, the support and confidence values were set at 0,2 and 0,7, respectively. In other words, rules are only interesting when the antecedent’s frequency is higher than 20% and the consequent is observed in 70% of the transactions.

$$GS = \kappa + \alpha_R Read + \alpha_C Cty + \alpha_E Eng + \alpha_Q Quant + \alpha_W Wrt \quad (4)$$

where:

- GS: global Score
- κ : intersect term
- α_R : reading skills score
- α_C : citizenship skills score
- α_E : English skills score
- α_Q : quantitative reasoning score
- α_W : writing skills score

$$\log \frac{p}{1-p} = \kappa + \alpha_W Women \quad (5)$$

where:

- p : depending on the analysis, p could be the probability of obtaining a high score (above the average) in any skill or the probability of obtaining a positive leap between the admission and Saber Pro tests (positive GLS).
- κ denotes an intersect term.
- α_W : $\log(OddsRatio)$ of being women. In other words, the ratio between the odds of women obtaining a high score or a positive GLS in any skill with respect to men.

Results

Descriptive statistics

Figure 4 shows the boxplot by gender on global scores in the Saber Pro test, where almost identical results are obtained by women and men (similar means and interquartile ranges). Consequently, no meaningful gap in global scores is observed. On the other hand, the evolution in GGS per skill since 2016 seems to be invariant (Figure 5), where the highest GGS are found in quantitative reasoning in favor of men (negative GGS) and writing communication in favor of women (positive GGS). In addition, this behavior is consistent on all campuses, as shown in Figure 6.

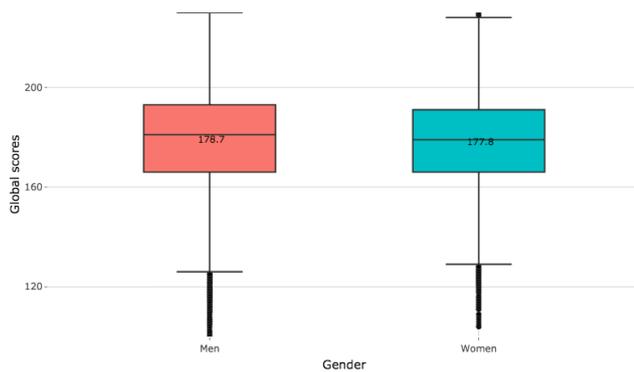


Figure 4. Gender gap on global scores
Source: Authors

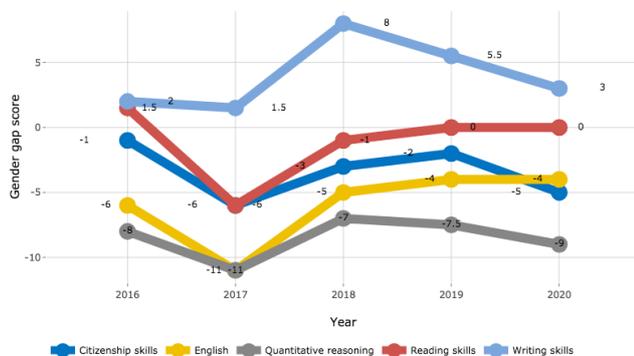


Figure 5. Evolution of gender gap scores per module
Source: Authors

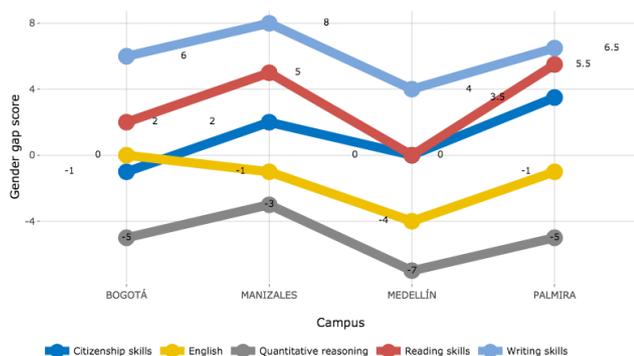


Figure 6. Evolution of gender gap scores per module per campus
Source: Authors

Table 1. GGS statistics per module (parametric approach)

Module	Shapiro-Wilks test, women (p-value)	Shapiro-Wilks test, men (p-value)	Fligner-Killeen homogeneity test (p-value)	T-test mean differences (p-value)	Effect size (D-Cohen)
Citizenship skills	2,07E-19	1,14E-19	1,15E-09	3,04E-03	Negligible
English	2,15E-18	1,70E-18	1,04E-01	1,52E-14	Negligible
Quantitative reasoning	1,04E-16	5,85E-24	1,08E-05	4,26E-44	Small
Reading skills	2,31E-10	1,01E-14	5,09E-05	9,62E-01	Negligible
Writing skills	4,70E-16	5,72E-19	3,70E-04	2,28E-08	Negligible

Source: Authors

Likewise, based on the non-parametric approach described above, Figure 7 and Table 2 show that the GGS per skill are well differentiated, given that the confidence intervals and interquartile ranges are very narrow. Therefore, the GGS behave very consistently. Citizenship skills, English, and quantitative reasoning are in favor of men, especially the latter. In contrast, written skills are in favor of women, whereas there is no gap in the reading comprehension skills.

These GGS results are coherent with some other findings in the international literature about gender-based gaps in academic performance at elementary and secondary schools, where men score higher than women in math but lower in reading tests (Abadía and Bernal, 2017). Consequently, our results suggest that the gender gap in math tests is still present in higher education, and even in the case of STEM programs, which is the main population addressed in this study.

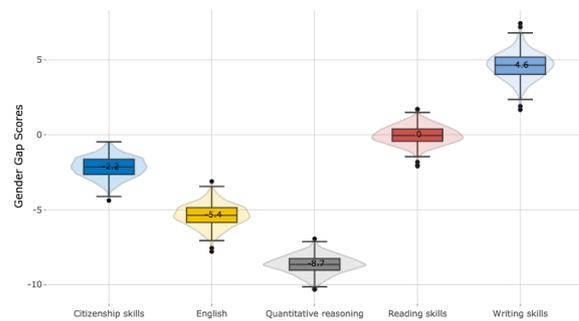


Figure 7. Gender gap per module (bootstrapping)
Source: Authors

In addition, the results shown in Table 1 suggest that the difference in means (between men and women) is statistically significant according to the p-values obtained for the statistical T-test. However, the D-cohen index also suggests that the effect size is almost negligible in most of the skills except for quantitative reasoning, where the gap is considered small.

Table 2. GGS statistics per module using a non-parametric bootstrapping approach - means and confidence intervals (95%)

Module	GGS (mean/conf. interval)
Citizenship skills	-2,2 / -3,4 to -0,8
English	-5,4 / -6,8 to -4
Quantitative reasoning	-8,7 / -9,8 to -7,5
Reading skills	0 / -1,2 to 1,1
Writing skills	4,6 / 3 to 6,3

Source: Authors

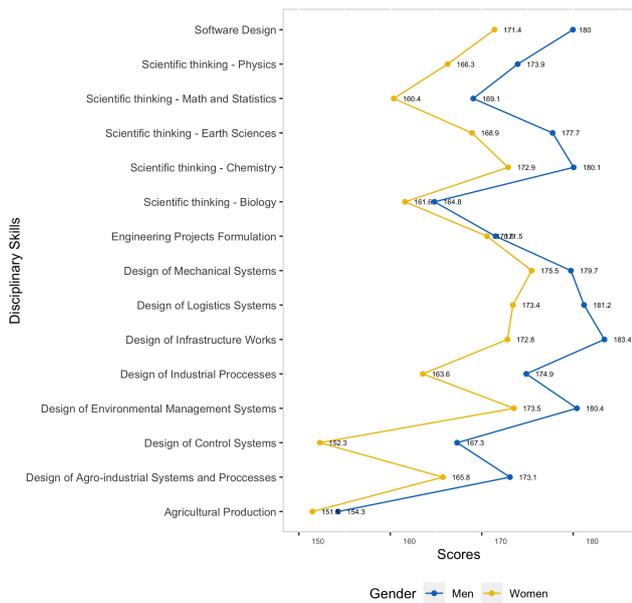


Figure 8. Gender gap for disciplinary skills
Source: Authors

Figure 8 shows the scores in each disciplinary skill by gender. As mentioned earlier, disciplinary skills are chosen by each curricular program. In general, the GGS are in favor of men in almost all disciplinary skills except for agricultural production, engineering project formulation, and scientific thinking in biology. The highest GGS is found in design of control systems, design of industrial processes, and design of infrastructure works.

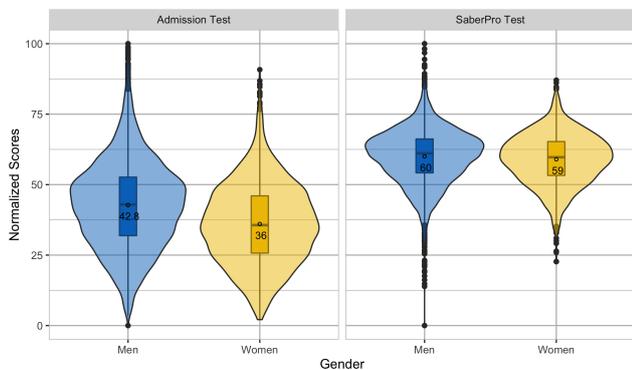


Figure 9. Escalated scores in the admission and Saber Pro tests by gender
Source: Authors

As for the GLS, Figure 9 shows the global escalated scores for both the admission and the Saber Pro tests, where there is a differentiated performance between men and women. Results suggest that the initial GGS gap in the admission test (42,8/36 points) is almost closed in the Saber Pro test (60/59). This behavior is confirmed in Figure 10, where women’s leap scores (mean: 22,9) are consistently higher than those of men (mean: 17,3).

Figures 11 and 12 show the GLS scores by social strata and schooling type. As far as social classes are concerned, the results suggest that not only is the leap higher for women in each class, but also that it is higher when the students come from the lower classes. In addition, the gender gap

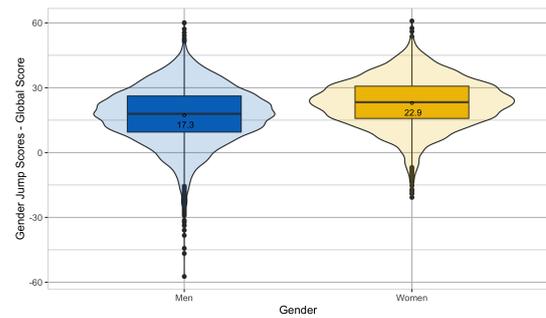


Figure 10. GLS for the global score
Source: Authors

is slightly higher in the lower strata. In terms of schooling type, the results show that the jump is consistently higher for students from public schools, with the gender gap being almost identical for both schooling types. These results are very promising regarding the mission of a public university, showing that not only do women experience the greatest jumps, but also that these jumps are more pronounced for women coming from public schools and lower classes.

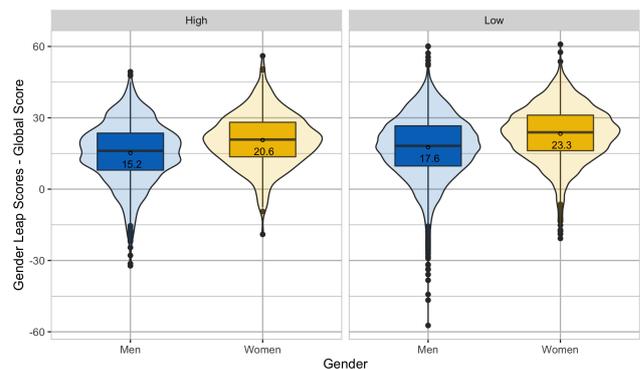


Figure 11. GLS for the global scores by social stratum
Source: Authors

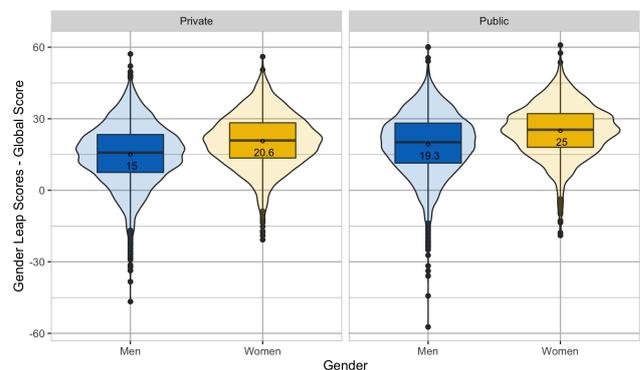


Figure 12. GLS for the global scores by schooling type
Source: Authors

Figure 13 shows the GLS by campus. In general, the greatest jumps are seen in the Palmira campus, where gender parity is significantly higher than on the other campuses. This finding is consistent with previous research showing that academic performance tends to improve in groups with higher proportions of women (Crombie et al., 2003).

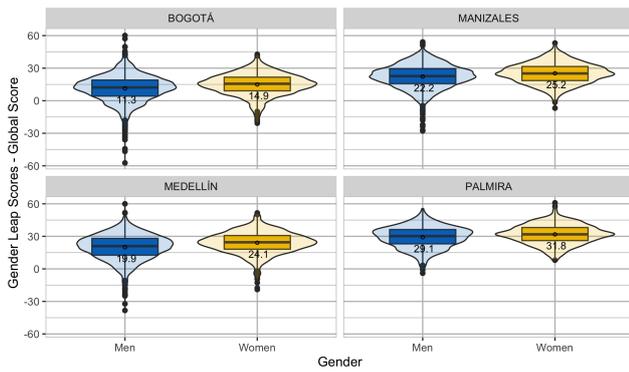


Figure 13. GLS for the global scores by campus
Source: Authors

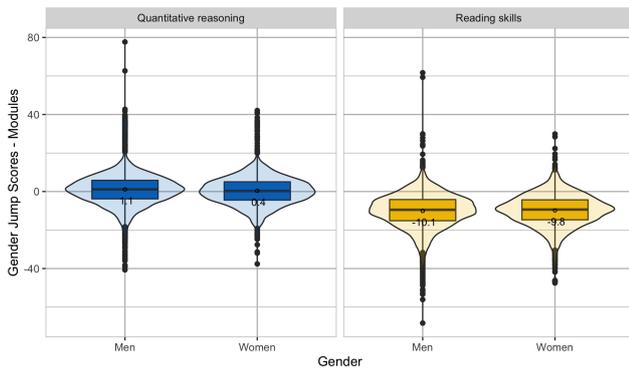


Figure 14. GLS for quantitative reasoning and reading skills
Source: Authors

In contrast, the GLS in quantitative reasoning and reading comprehension are very similar for both genders (Figure 14). Regarding quantitative reasoning, there is no change between men and women, so the gender gap is not only almost identical, but also academic performance remains invariant (GLS close to zero) in the Saber Pro test when compared to the admission exam. On the contrary, the performance in reading comprehension tends to deteriorate along the university experience, as the GLS are negative (GLS close to -10 for both genders). However, performance deteriorates in the same proportion in men and women. These results are confirmed by the statistical tests shown in Table 3, where, for most of the skills, the means for the GLS are statistically different between men and women [§]. However, these differences are negligible for quantitative reasoning and reading skills, as well as small for the global score (D-Cohen result).

Regression analysis

Correlograms

Figures 15 and 16 show correlograms for the escalated scores in each module of the admission test (suffix: Adm) and the Saber Pro Test (suffix: Spro) for men and women. In general, there are moderate associations between the admission and Saber Pro global scores. Therefore, a higher/lower admission score does not necessarily imply higher/lower scores in the exit exam (Saber Pro). Moreover, this association is a bit stronger for women. In other

Table 3. Statistics for the GLS

Module	Shapiro-Wilks test, women (p-value)	Shapiro-Wilks test, men (p-value)	Fligner-Killeen homogeneity test (p-value)	T-test mean differences (p-value)	Effect Size (D-Cohen)
Quantitative reasoning	2E-14	1E-18	1E-01	1E-04	Negligible
Reading skills	4E-07	3E-11	8E-03	0,16	Negligible
Global score	2E-04	1E-08	9E-12	3E-82	Small

Source: Authors

words, there is an effect of the university experience on the academic performance in the exit exam.

On the other hand, reading, citizenship and English (second language) skills are the scores that better correlate with the global scores in the Saber Pro test. Surprisingly, quantitative reasoning is not as strongly correlated with the final global score as reading or English skills, even in the case of engineering students. That is to say that, instead of quantitative reasoning, now reading, citizenship, and English are the competencies that make the difference in the global score. In addition, the correlation coefficients (R) remain almost invariant between men and women, as shown in Table 4.

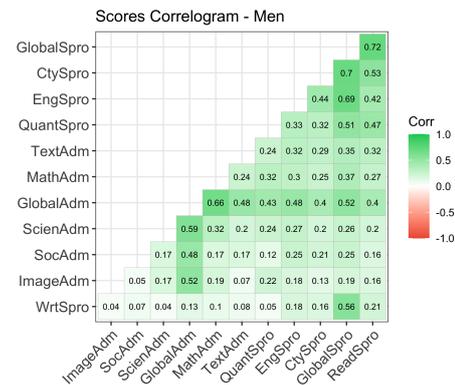


Figure 15. Correlogram for the admission and exit exam scores (men)
Source: Authors

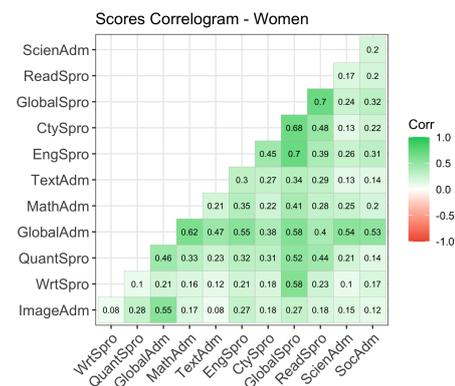


Figure 16. Correlogram for the admission and exit exam scores (women)
Source: Authors

[§]In the case of reading comprehension skills, the p-value is non-conclusive, so there is no statistical evidence that the means for the GLS are different between men and women. This may be explained by the fact that both means are very close, i.e., -10,1 and -9,8 for men and women, respectively.

Table 4. Correlation coefficients (R) for the global score in each skill per gender

Skill	R Men	R Women
Reading	0,72	0,7
Citizenship	0,7	0,68
English	0,69	0,7
Quantitative reasoning	0,51	0,52
Writing	0,56	0,58

Source: Authors

Linear multivariate regressions

As mentioned above, some linear regressions on the Saber Pro global scores per gender were modeled following Equation (4). A linear model was chosen based on the AIC (Akaike Information Criterion) using a step-wise algorithm. The model started with both the admission and Saber Pro module scores. However, the admission scores were neglected, since no significant changes were observed regarding Pearson’s R^2 correlation when they were included, going from 0,921 to 0,923. This correlation value not only indicates that 92% of the variance for the global score is explained by the skill scores, but also that this result is coherent with the poor correlation coefficients shown in Figures 15 and 16. Again, the results suggest that there is an effect of the university experience on academic performance in the exit exam.

Table 5 shows the resulting regression coefficients for men and women following Equation (4). In both cases, R^2 is high enough to explain most of the global score variance (0,922 and 0,919 for men and women, respectively), with all the coefficients being statistically significant ($p_{values} < 2e - 16$ for a T-test in both cases) and standard errors of 0,03314 and 0,0325 for men and women, respectively. The regression coefficients remain almost invariant for both genders, and English and citizenship skills are the modules that mostly explain the global scores variance, whereas quantitative reasoning is the skill that most poorly explains it. Moreover, English explains this variance slightly better when the students are women, just as Citizenship skills when the students are men.

Table 5. Regression coefficients by gender, Equation 4

Gender	κ	α_R	α_C	α_E	α_Q	α_W
Men	-0,258	0,294	0,346	0,372	0,166	0,328
Women	-0,269	0,306	0,327	0,395	0,163	0,327

Source: Authors

Logistic regressions

A set of logistic regressions was modeled to estimate the odds ratios between men and women for different skill scores. In general, the modeled logistic regressions followed the formulation shown in Equation 5. However, several sets of regressions were tested for different purposes, as follows:

1. A first set determined the odds ratios of obtaining "high" scores (defined as scores above the average) for any general skill.
2. A second set determined the odds ratios of obtaining "positive" GLS.

3. A third set determined the odds ratios of obtaining higher scores in disciplinary skills when compared to general ones.

Table 6 shows the results for the odd ratios of obtaining a "high" score in any skill. Most of the odds ratios are less than 1, implying that, for most of the skills, a woman’s odds of obtaining a high score are only a portion of men’s odds. The most critical difference is observed in quantitative reasoning, where a woman’s odds of getting a high score are only 55% of men’s odds. In other words, men’s odds of getting a high score in quantitative reasoning almost duplicate those of women. On the contrary, in the case of writing skills, women’s odds of getting a high score are 1,27 times higher than those of men. In addition, most of the obtained odds ratios are statistically significant, except for the case of reading comprehension, where no conclusions can be drawn, as there is no distinction between the odds of women and men, *i.e.*, very close to 1 (0,94).

Table 6. Statistics for the odds ratios of getting a high score in any general skill

Skill	α_W^a	OddsRatio	p-value
Global score	-0,249	0,78	7,40E-08
Reading	-0,056	0,94	0,2222
Citizenship	-0,140	0,87	0,00269
English	-0,360	0,70	1,40E-14
Quantitative reasoning	-0,59	0,55	2,00E-16
Writing skills	0,237	1,27	4,20E-07

Source: Authors

$$^a(\ln(\text{OddsRatio}))$$

Table 7 shows the results for the odd ratios of obtaining a "positive" GLS score in any skill. Again, a woman’s odds of getting a positive GLS score in quantitative reasoning are less than those of men (82%). On the contrary, in the case of the global score, women’s odds of getting a positive GLS are 3,53 times higher than men’s. Moreover, most of the obtained odds ratios are statistically significant, except for the case of reading comprehension, where no conclusions can be drawn, as there is no distinction between the odds of women and men, *i.e.*, very close to 1 (0,987).

$$\ln\left(\frac{P(\text{Score}_{disc} > \text{Score}_{gen})}{1 - P(\text{Score}_{disc} > \text{Score}_{gen})}\right) = -0,3679 - 0,4624 * \text{Women} \quad (6)$$

Finally, Equation 6 shows the final regression coefficients for the odds of getting a higher score in disciplinary skills when compared to general ones. If the students are men, the odds are only 44%, which means that general skill scores are very often higher than disciplinary ones. However, if students are women, the odds decrease down to 14%, which is not only very low but also means that the odds of getting a higher score in disciplinary skills for a woman are much less than those of men (34%). This odds ratio is statistically significant, as shown in Table 8.

Association Rules

This section presents the results for the data mining analysis based on association rules ($X \Rightarrow Y$) using the *Apriori* algorithm, as described in the *Methodology* section. The analysis followed two main approaches:

Table 7. Statistics for the odds ratios of getting a positive GLS per skill

Skill	α_W^a	OddsRatio	p-value
Quantitative reasoning	-0,192	0,82	3,8E-05
Reading	-0,012	0,98	0,87
Global score	1,261	3,53	2E-16

Source: Authors

^a(Ln(OddsRatio))

Table 8. Statistics for the odds ratios of getting a higher score in disciplinary skills when compared to general ones

Event	α_W^a	OddsRatio	p-value
$P(\text{Score}_{disc} > \text{Score}_{gen})$	-0,4624	0,34	2E-16

Source: Authors

^a(Ln(OddsRatio))

- Based on the skills with the highest GGS (quantitative reasoning, English, and writing), association rules were mined, with gender as the rule’s consequent. The resulting mined rules are shown in Table 9.
- The exogenous variables of social Stratum (high/low) and schooling type (public/private) were included, and association rules were separately mined for both men and women. A comparison of the support and confidence values obtained is presented for the common mined rules, in order to determine the cases where these rules are stronger. The resulting mined rules are shown in Table 10.

Table 9 confirms some of the results shown in the previous sections. In general, men scoring "high" in quantitative and English but "low" in writing skills is a strong rule, with a confidence greater than 75%. This result is consistent with the high GGS in favor of men for these skills, as was previously analyzed. The fact that men tend to score simultaneously high in the couples [quantitative, English] and [English, writing] but also simultaneously low in the couple [quantitative, writing] were also detected as strong rules. This result is very interesting, as it implies a certain type of association in male engineering students, suggesting that a good performance in quantitative reasoning comes along with a poor performance in writing skills.

Table 10 shows a comparison of the support and confidence values for men and women with regard to the separately mined rules. First of all, the confidence values are almost identical. Therefore, the mined rules are equally confident between men and women. However, the rules are not necessarily equally stronger, as the support values exhibit different values for both genders. As previously mentioned, the social stratum and schooling type were explicitly included in the mined rules with the aim to assess their effect on skill scores. The main results may be stated as follows:

1. Coming from lower classes,
 - Men engineering students score "high" more often than women in quantitative reasoning.

- Women engineering students score "low" more often than men in second language skills (English). This result is very interesting, given that some other reported works (Corpas, 2013) have found that women score higher in English as a second language. This is not the case for engineering students.
- Women engineering students score simultaneously "low" more often than men in quantitative reasoning and second language skills (English).

2. Coming from public schools,

- Women engineering students score simultaneously "low" more often than men in quantitative reasoning and second language skills (English).

3. Coming from lower classes and public schools,

- Women engineering students score "low" more often than men in quantitative reasoning.

Table 9. Obtained association rules ($X \Rightarrow Y$) using the Apriori algorithm: support (Sup) >0,2 and confidence (Conf) >0,7

Antecedent (X)	Consequent (Y)	Sup	Conf
Quant=High	Sex=Men	0,39	0,78
Eng=High	Sex=Men	0,39	0,76
Wrt=Low	Sex=Men	0,37	0,75
Quant=High & Eng=High	Sex=Men	0,25	0,79
Quant=High & Wrt=Low	Sex=Men	0,20	0,80
Eng=High & Wrt=High	Sex=Men	0,21	0,79

Source: Authors

Table 10. Comparison of support and confidence values for men and women in the obtained association rules ($X \Rightarrow Y$) applying the Apriori algorithm (support >0,2 and confidence >0,7)

Ant ^a (X) \Rightarrow	Cons ^b (Y)	Sup (M/W) ^c	Con(M/W) ^d
(Quant=High)	Strat ^e =Low	0,46/0,33	0,84/0,82
(Eng=Low)	Strat=Low	0,43/0,52	0,93/0,94
(ScTyp ^f =Public & Quant=Low)	Strat=Low	0,27/0,35	0,97/0,98
(Quant=Low & Eng=Low)	Strat=Low	0,26/0,37	0,93/0,95
(ScTyp=Public & Quant=Low)	Eng=Low	0,20/0,29	0,72/0,8
(Quant=Low & Eng=Low)	ScTyp=Public	0,2/0,29	0,74/0,73

Source: Authors

^aRule’s antecedent.
^bRule’s consequent.
^cRule’s support (men/women).
^dRule’s Confidence (Men/Women).
^eSocial stratum
^fSchooling type.

Conclusions

Gender inequalities in academic performance at elementary and secondary schools have been widely reported, evidencing that the lack of skills in mathematics and reading may lead to fewer chances of being admitted to STEM programs. In particular, Abadía and Bernal (2017) and Abadía (2017) specifically found that men score better globally and in math and science, as do women in reading, in the high school exit examination known as *Saber 11*. Surprisingly, the results obtained in this study suggest that

the gender gap in math tests is still present in higher education, and more specifically in the addressed STEM programs, which are expected to be math-oriented. One of the possible explanations, regardless of the mathematics background, arises from psychological factors such as women's anxiety, which has been previously reported in standardized tests (De la Rica and Rebollo, 2018). In addition, not only quantitative reasoning exhibited a gender gap, but also other skills including disciplinary competencies.

The highest gender gap score (GGS) in favor of men was found in quantitative reasoning. According to the mathematical analysis performed, this gap is statistically significant, but its effect is small based on the gender score variances (D-cohen). Moreover, there is no improvement from the admission exam scores, since the gender gap remains almost identical in the college exit exam (Saber Pro). In addition, men's odds of getting a high score in quantitative reasoning almost duplicate those of women. On the other hand, the data mining approach is also consistent, as it confirms that men scoring high in quantitative and English are strong rules, with a confidence greater than 75%. Moreover, another strong rule shows that, when students come from lower classes or public schools, women score low more often than men.

In contrast to the results reported in Abadía and Bernal (2017), where men score better than women in high school reading tests, college exit exams show no statistically significant gender gap for engineering students. This may lead to the idea that the gender gap has been closed. However, our results show that reading not only experienced the highest leap from admission tests, but also the most negative, which means that there is a significant weakening in reading skills. This result may lead to deep thoughts about the reading comprehension capabilities in STEM curricular programs that, apparently, are no longer developed since high school.

Another result reported in (Abadía and Bernal, 2017) shows that men score better than women in high school global scores. This is consistent with our findings with regard to the admission tests, where a gender gap in favor of men was also observed. However, this gap is completely closed in the college exit exam, implying that women improve more than men as a result of the college experience. The results suggest that women's odds of improving their admission scores are 3,53 times higher than those of men. In addition, the models for global scores show that the regression coefficients remain almost invariant for both genders where English and citizenship skills are the most significant. In other words, quantitative reasoning is not a determining factor in the global scores of engineering students, since it poorly explains the variances. Consequently, the closing of the gender gap in global scores is not necessarily related to quantitative reasoning or reading skills.

In terms of the GLS, the results show that females close the gap in the admissions test, which entails higher score jumps. Furthermore, the highest score jumps are observed on the campus where gender parity is significantly higher (Palmira). Additionally, not only do women experience the larger jumps, but these jumps are more pronounced for women from public schools and lower-class backgrounds.

The highest GGS in favor of women was found in writing skills. Mathematically speaking, this gap is also statistically significant, but its effect is considered smaller when compared to the gap in quantitative reasoning. In addition, women's odds of getting a high score in writing skills are 1,27 times higher than those of men. The data mining approach was also consistent, confirming that men scoring low in writing skills is a strong rule, with a confidence greater than 75%. However, another mined association rule showed a very interesting result, implying that a good performance in quantitative reasoning for male students comes with a poor performance in writing skills.

Finally, regarding disciplinary competencies, the results show a significant gender gap in favor of men. This result is very consistent in most of the disciplinary areas of engineering. Moreover, men's odds of getting a higher score in disciplinary skills are only 44% when compared to general skills (quantitative reasoning, reading, writing, etc.), which means that general skill scores are very often higher than those of disciplinary skills. In addition, for women students, the odds decrease down to 14%, which is not only very low but also means that the odds of getting a higher score in disciplinary skills for a woman are much less than those of men (34%). A possible explanation for this result may be the reported idea that women's performance is higher at tests that assess knowledge acquisition rather than knowledge application (UNESCO, 2019). In any case, this result may lead to deep thoughts about disciplinary courses in STEM programs, which probably not only need to be improved but also might have to be gender-oriented.

CRediT author statement

All authors: conceptualization, methodology, software, validation, formal analysis, investigation, writing (original draft, writing, re-view, and editing), data curation.

References

- Alam, A. (2022, April). Psychological, sociocultural, and biological elucidations for gender gap in STEM education: a call for translation of research into evidence-based interventions. In Alam, A. *Psychological, Sociocultural, and Biological Elucidations for Gender Gap in STEM Education: A Call for Translation of Research into Evidence-Based Interventions. Proceedings of the 2nd International Conference on Sustainability and Equity (ICSE-2021). Atlantis Highlights in Social S.* <https://ssrn.com/abstract=4073986>
- Abadía, L K., and Bernal, G. (2017). A widening gap? A gender-aased analysis of performance on the Colombian high school exit examination. *Revista de Economía del Rosario*, 20(1), 5-31. <https://doi.org/10.12804/revistas.urosario.edu.co/economia/a.6144>
- Abadía, L. (2017). Gender score gaps of Colombia in PISA test. *Vniversitas Económicas*, 17(8), 1-24.
- Brenøe, A. A., and Zolitz, U. (2020). Exposure to more female peers widens the gender gap in stem participation. *Journal of Labor Economics*, 38(4), 1009-1054. <https://doi.org/10.1086/706646>.
- Cárcamo, C., and Mola, J. (2019). Diferencias por sexo en el desempeño académico en Colombia: Un análisis regional.

- Economía & Región*, 6(1), 133-169. <https://revistas.utb.edu.co/index.php/economia-region/article/view/137>
- Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias, *The Quarterly Journal of Economics*, 134(3), 1163-1224. <https://doi.org/10.1093/qje/qjz008>
- Corpas, M. (2013). Gender differences in reading comprehension achievement in English as a foreign language in Compulsory Secondary Education. *Didactics of Language and Literature. Education*, 17, 67-84. <https://tejuelo.unex.es/article/view/2546>
- CPEM, DANE, and ONU (2020). *Mujeres y hombres: Brechas de género en Colombia*. Consejería Presidencial para la Equidad de la Mujer. <https://www.dane.gov.co/files/investigaciones/genero/publicaciones/mujeres-y-hombre-brechas-de-genero-colombia-info\rme.pdf>
- Crombie, G., Pyke, S.W., Silverthorn, N., Jones, A., and Piccinin, S (2020). Students' perceptions of their classroom participation and instructor as a function of gender and context. *The Journal of Higher Education* 74(1), 51-76. <https://doi.org/10.1353/jhe.2003.0001>
- De la Rica, S., and Rebollo-Sanz, Y. F. (2018). Brechas de género en competencias cognitivas: evidencia internacional. *Cuadernos Económicos De ICE*, 95, 125-149. <https://doi.org/10.32796/CICE.2018.95.6645>
- Gallego, L. (2021). *Salto y brechas*. Universidad Nacional de Colombia. <https://repositorio.unal.edu.co/handle/unal/80700>
- García-Holgado, A., Mena, J., García-Peñalvo, F. J., Pascual, J., Heikkinen, M., Harmoinen, S., García-Ramos, L., Peñabaena-Niebles, R., and Amores, L. (2020). Gender equality in STEM programs: a proposal to analyse the situation of a university about the gender gap. In IEEE (Eds.), *2020 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1824-1830). IEEE. <https://doi.org/10.1109/EDUCON45650.2020.9125326>
- García-Holgado, A., Camacho, A., and García-Peñalvo, F. J. (2019). La brecha de género en el sector STEM en América Latina: Una propuesta europea. In M. L. Sein-Echaluce Lacleta, Á. Fidalgo-Blanco, and F. J. García-Peñalvo (Eds.), *Actas del V Congreso Internacional sobre Aprendizaje, Innovación y Competitividad - CINAIC 2019* (pp. 704-709). Servicio de Publicaciones Universidad de Zaragoza. <http://dx.doi.org/10.26754/CINAIC.2019.0143>
- Harris, D., Lowrie, T., Logan, T., and Hegarty, M. (2021). Reasoning, mathematics, and gender: Do spatial constructs differ in their contribution to performance?. *The British Journal of Educational Psychology*, 91(1), 409-441. <https://doi.org/10.1111/bjep.12371>
- Jungert, T., Hubbard, K., Dedic, H., and Rosenfield, S. (2019). Systemizing and the gender gap: examining academic achievement and perseverance in STEM. *European Journal of Psychology of Education*, 34(2), 479-500. <https://doi.org/10.1007/s10212-018-0390-0>
- Kuschel, K., Ettl, K., Díaz-García, C., and Alsos, G. A. (2020). Stemming the gender gap in STEM entrepreneurship - insights into women's entrepreneurship in science, technology, engineering and mathematics. *International Entrepreneurship and Management Journal*, 16, 1-15. <https://doi.org/10.1007/s11365-020-00642-5>
- ICFES (2021). *Guías de Orientación Saber Pro 2021 - Módulos de Competencias Genéricas - Bogotá Colombia*. <https://www.icfes.gov.co/documents/20143/2297330/Guia+de+orientacion+modulos+genericos+Saber-Pro+2021.pdf>
- Kucian, K., Loenneker, T., Dietrich, T., Martin, E., and Von Aster, M. (2005). Gender differences in brain activation patterns during mental rotation and number related cognitive tasks. *Psychology Science*, 47(1), 112-131.
- Lawton, C. A., and Hatcher, D. W. (2005). Gender differences in integration of images in visuospatial memory. *Sex Roles*, 53(9), 717-725. <https://doi.org/10.1007/s11199-005-7736-1>
- Makarova, E., Aeschlimann, B., and Herzog, W. (2019, July). The gender gap in STEM fields: The impact of the gender stereotype of math and science on secondary students' career aspirations. In *Frontiers in Education* (vol. 4, p. 60). Frontiers Media SA. <https://doi.org/10.3389/feduc.2019.00060>
- MEN and OECD (2016). *La educación en Colombia*. https://www.oecd-ilibrary.org/education/education-in-colombia_9789264250604-en
- Raabe, I. J., Boda, Z., and Stadtfeld, C. (2019). The social pipeline: How friend influence and peer exposure widen the STEM gender gap. *Sociology of Education*, 92(2), 105-123. <https://doi.org/10.1177/0038040718824095>
- Radinger, T., Echazarra, A., Guerrero, G., and Valenzuela, J. P. (2018). *OECD reviews of school resources: Colombia*. <https://www.oecd.org/education/school/OECD-Reviews-School-Resources-Summary-Colombia-Spanish.pdf>
- Reinking, A., and Martin B. (2018). The gender gap in STEM fields: Theories, movements, and ideas to engage girls in STEM. *Journal of New Approaches in Educational Research*, 7(2), 148-153. <https://doi.org/10.7821/naer.2018.7.271>
- Rincón-Baez, W. U., and Arias-Velandia, N. (2019). Brecha de rendimiento académico por género en Saber Pro en programas de Administración de los departamentos de Colombia. *Panorama*, 13(25), 142-161. <https://doi.org/10.15765/pnrm.v13i25.1410>
- Reilly, D., Neumann, D. L., and Andrews, G. (2017). Gender differences in spatial ability: Implications for STEM education and approaches to reducing the gender gap for parents and educators. In *Visual-Spatial Ability in STEM Education* (pp. 195-224). Springer. https://doi.org/10.1007/978-3-319-44385-0_10
- Riegle-Crumb, C. (2005). *The cross-national context of the gender gap in math and science*. In L. V. Hedges and B. Schneider (Eds.) *The Social Organization of Schooling* (pp. 227-243). Russell Sage Foundation.
- Tandrayen-Ragoobur, V., and Gokulsing, D. (2021). Gender gap in STEM education and career choices: what matters?. *Journal of Applied Research in Higher Education*. <https://doi.org/10.1108/JARHE-09-2019-0235>
- UNESCO (2019). *Descifrar el código: la educación de las niñas y las mujeres en ciencias, tecnología, ingeniería y matemáticas (STEM)*. <https://unesdoc.unesco.org/ark:/48223/pf0000366649>.

Verdugo-Castro, M., Sánchez-Gómez, C., and García-Holgado, A. (2018, October 24-26). *Gender gap in the STEM sector in pre and university studies of Europe associated with ethnic factors*. In Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'18). Association for Computing Machinery, New York, NY, USA, 984-990. <https://doi.org/10.1145/3284179.3284348>

World Economic Forum (WEF) (2020). *Global gender gap report 2020*. https://www3.weforum.org/docs/WEF_GGGR_2020.pdf