# THE RIGHT TO EDUCATION AND PERFORMANCE GAPS IN THE SABER PRO TESTS IN COLOMBIA: PERSPECTIVES FROM SOCIOECONOMIC AND ACADEMIC FACTORS.

Gustavo García-Cediel[a], Diego Cardona-Arbelaez[b], Carlos Cely-Charris[c]

[1]PhD.,Doctor of Business Administration, professor at the Faculty of Economic Sciences at the University of Cartagena, Cartagena (Colombia)
[2]PhD.,Doctor of Administration, professor at the Faculty of Economic Sciences at the University of Cartagena, Cartagena (Colombia).
[3]Mag., MBA in Marketing. professor at the Faculty of Economic Sciences at the University Libre of Cartagena, sede Cartagena (Colombia).

ggarciac1@unicartagena.edu.co.[1]
dcardonaa@unicartagena.edu.co[2]
carlosalbertocely@hotmail.com[3]

**Abstract**

The main objective of this study was to identify the sociodemographic, economic, and institutional factors that influence student performance on the Saber Pro tests in Colombia. The research was conducted using a quantitative, descriptive, and correlational approach with a non-experimental design. The Ordinary Least Squares (OLS) regression technique was applied to a database comprising 139,288 students who took the tests during the 2023-1 and 2024-1 periods. The variables analyzed included gender, socioeconomic status, age, educational modality (face-to-face or virtual), parents' educational level, internet access, method of preparation before the test, form of payment for studies, and number of hours worked. Among the most relevant findings is that men obtain, on average, better results than women, although the difference is slight. Socioeconomic status shows a positive relationship with scores: students in strata 4, 5, and 6 obtain significantly higher scores. Likewise, students enrolled in face-to-face programs perform better than those studying in distance learning programs. Internet access and the fact that parents have postgraduate education are also associated with better performance. In terms of preparation methods, those who took formal courses obtained better results than those who did not prepare or only reviewed on their own. Based on these results, it can be concluded that there are structural inequalities that affect academic performance. It is recommended that public policies be implemented that focus on reducing gender and socioeconomic gaps starting in basic education, expanding access to technological resources, and strengthening educational quality in public institutions. Finally, it is suggested that longitudinal studies be conducted to observe the evolution of student performance throughout the educational process.

**Keywords**: Saber Pro, higher education, socioeconomic factors, academic performance, public policies.

## 1. Introduction

Human Capital Theory (Becker, 1964) posits that the accumulation of knowledge and skills through education and training is an investment that increases individuals' productivity and future income. Therefore, in the context of the Saber Pro tests, this theory is particularly relevant, as the results obtained by students reflect the level of human capital acquired during their university education and could therefore be considered the appropriate theoretical framework to support this research. The ability of graduates to apply generic and specific skills in the workplace depends

largely on the quality of the education received at universities, as well as on the investment that each student has made in their own learning.

Human capital theory highlights the importance of specific and general training. Within this framework, the Saber Pro tests assess both generic skills, applicable to multiple work environments, and specific skills related to the area of study. The results obtained not only allow universities to improve their curriculum, but also provide students with valuable information about their level of competitiveness in the market.

The Saber Pro tests are a fundamental tool for evaluating the quality of higher education for students in Colombia, measuring the generic and specific competencies of students nearing graduation, which apply to both undergraduate students (Saber 11) and university students about to graduate. The results of this test allow higher education institutions to formulate actions to increase their strengths and overcome their weaknesses by taking actions such as adjusting their curriculum and/or changes to their teaching activities.

Despite the significant usefulness of an analysis of the results of these tests for the design of public policies, there is no detailed analysis or systematic monitoring of the results of the Saber Pro tests for the majority of students who take them at the university level in literature. This document seeks to fill this gap and analyze the results of the Saber Pro tests for all programs, asking whether certain sociodemographic, economic, and institutional variables influence the results.

In this vein, the research question could be posed as follows:

What are the determining factors that explain the results of the Saber Pro tests, considering the demographic, sociodemographic, economic, and institutional characteristics of the students?

This study seeks to generate relevant information for academic decision-making and improving educational quality at the institution. This first section presents the problem statement, the objective, and the justification for the study. The following section reviews the preceding literature, and finally presents the methodology, results, discussion, and policy implications.

## 2. Literature review

Given that in order to take the Saber Pro tests, students must fill out data on sociodemographic, economic, and institutional aspects, this dataset has become an important input when trying to understand their results. The main studies that have been developed on this topic are outlined below.

In this vein, Cifuentes (2013) aimed to predict the performance of students on the SABER PRO standardized tests for the Economics program based on the admission process for selecting candidates. The "r" method used in the research was Quantile Regression, a statistical technique proposed by Koenker and Basset (1978) that allows estimating relationships between variables in different quantiles of the distribution of the dependent variable, rather than focusing solely on the conditional mean as in ordinary least squares. This approach is especially useful in the presence of asymmetry in the response variable, outliers, and when the assumptions of normality are not met. The result of this work is given through least squares regression with robust errors, as it shows that all the variables included are significant at 5%. A positive effect of male gender on the prediction of the SABER PRO test score is highlighted, indicating that men tend to obtain better results. In addition, the coefficients for mathematics, chemistry, and history have a greater impact on the

prediction. Age has a negative effect, suggesting that younger students at the time of taking the SABER 11 test achieve better scores on the SABER PRO test.

The study by Osma et al. (2014) focused on analyzing the factors that influence the results of the 2010 Saber Pro tests among students in their final semesters of civil engineering in Colombia, with the aim of contributing to policies for improving educational quality in higher education institutions. Using SPSS, Excel, and SPAD, it was found that the variables of gender, type of education, parents' educational level, and socioeconomic status significantly influence the scores obtained, while work and degree are not related. The results show that above-average scores correspond to men from public universities belonging to the middle and upper socioeconomic strata.

On the other hand, the author Bahamón (2014) explores the factors that influence student performance on the Saber Pro tests, using a group of 68 psychology students in Colombia as a case study. Aspects related to intellectual capacity; sociodemographic and academic factors are highlighted. The results suggest that characteristics such as previous good academic performance, interests aligned with the area of study, and effective study habits are linked to better test performance. Likewise, significant differences in performance are observed according to variables such as socioeconomic context and type of educational institution.

On the other hand, Ramírez (2014) analyzes the socioeconomic and educational factors associated with the academic performance of students who took the Saber Pro-2009 test in Colombia, differentiating by gender and level of education (technical, technological, and professional). The results highlight that previous academic performance, as measured by the Saber 11 tests, has a strong and positive association with performance in higher education, while socioeconomic and educational variables have a weak association. It also concludes that the type of institution (university or non-university) and the characteristics of the programs influence academic performance in different ways.

For their part, Melo-Becerra, Ramos & Hernández-Santamaría (2017) examine the situation of higher education in the country and evaluate the efficiency levels of various academic programs using stochastic frontier techniques, based on results from the Saber Pro tests. The findings highlight that, although institutions can improve in efficiency, students' socioeconomic factors, such as family income and parental education, are key determinants of results. The study also underscores the importance of institutional variables such as infrastructure and teaching quality.

Another study on this topic was conducted by Acero (2017), who analyzed the demographic, socioeconomic, and academic factors that influence the results of the generic component of the Saber Pro tests in electromechanical engineering students at the UPTC, Duitama campus. The study, with a descriptive and inferential quantitative approach, applied tests of difference of means, correlations, and models such as GAMLSS and multinomial ordinal, processed with R software (version 3.4.1). The results show that socioeconomic status does not generate significant differences in most components, except in English and quantitative reasoning, wher s do have an influence. Excellent performance in quantitative reasoning and average performance in English stand out, confirming previous studies on the university. In addition, maternal occupation, especially when she is a laborer or housewife, is positively related to scores in quantitative reasoning and critical reading.

Vásquez (2018) sets as his main objective to determine whether the scores obtained in the SABER 11th grade test can be considered a valid predictor of performance in the SABER PRO test. To this end, he adopts a quantitative methodology based on data collection, statistical analysis, and hypothesis testing, which allows him to establish causal relationships between variables. Through numerical measurements and statistical tests, Vásquez Arrieta identifies certain patterns of behavior, providing educational institutions with relevant information for decision-making and improving their academic plans. However, the study also highlights that, although there is some correlation, there is no conclusive relationship that allows us to definitively state that the results of the SABER 11th grade test predict performance on the SABER PRO, which leaves open the need to continue investigating other variables with potential influence on academic performance.

Rincón-Báez & Arias (2019) examine the differences in academic performance between men and women in the SABER PRO tests during 2016-2018. They use a gender performance gap index to measure these differences at the departmental and national levels in Colombia, focusing on five key competencies: quantitative reasoning, critical reading, civic competencies, English, and written communication. The results highlight persistent gaps in favor of men in most competencies, except in written communication, where women perform better. Geographical variations linked to structural inequalities and regional opportunities are also evident.

Timarán-Pereira et al. (2020) aim to analyze the academic performance of engineering students in the SABER PRO tests, using educational data mining techniques. Through this approach, the authors seek to identify the socioeconomic, academic, and institutional factors that significantly influence the results of these assessments. The research not only contributes to the understanding of academic performance in the context of higher education in Colombia, but also provides key information that can guide decision-making in educational institutions, thus promoting the design of more effective and equitable pedagogical strategies. The research was descriptive in nature with a quantitative approach and a non-experimental design, which identifies patterns of academic performance. In addition, the CRISP-DM methodology, one of the most widely used in data mining, was employed. This methodology consists of six phases: problem analysis, data understanding and preparation, modeling, evaluation, and implementation, allowing for a structured approach to the analysis of results. The analysis of academic performance in the Saber Pro tests of engineering students in Colombia shows that institutional accreditation is a key factor for good performance, which is consistent with previous studies. In addition, the mode of study has a significant influence: students in face-to-face programs perform better than those in distance learning programs. It was also found that socioeconomic status impacts performance, as students from high and middle socioeconomic strata obtain better scores than those from low socioeconomic strata. These findings reinforce the relationship between educational quality, mode of study, and socioeconomic conditions with academic performance.

Narváez (2022) designed a statistical model that determines the academic factors affecting test results. The methodology used was a study based on multivariate and learning techniques, with the aim of establishing a connection between academic and sociodemographic variables and their influence on the results of the Saber Pro test. To this end, a multivariate statistical model was designed and selected to optimally identify the academic factors that affect test scores. No significant differences were found between the models and the reality reflected in the validation sample, with the exception of the PCME test, where the Random Forest model did not ly test the validation hypotheses. It was observed that the multivariate linear regression model does not show

significant differences in any of the tests, while Random Forest does show differences for certain values in ING and FPI, in addition to rejecting hypotheses of equality in the PCME test.

Demarchi (2023) focuses on understanding the factors that affect academic performance in this case, the Saber Pro state tests. A qualitative approach was used to identify the factors that affect student performance on the Saber Pro state tests. To organize and analyze the information, Atlas-Ti qualitative research software was used. In the case of interviews, they were transcribed and then analyzed in the software. It was concluded that performance on the Saber Pro tests is influenced by multiple factors, including the family environment, employment status, economic condition, and the neighborhood where the student lives. These elements can significantly affect the results obtained on the tests.

Canova-Barrios et al. (2023) aim to analyze the relationship between various personal, family, and institutional variables and the performance of students in the Nursing program at the University of Magdalena (Colombia) on the SABER PRO tests during the period between 2016 and 2019. This research seeks to identify patterns and factors that significantly affect students' academic results, considering aspects such as socioeconomic environment, family support, institutional conditions, and individual student characteristics. A descriptive and correlational study was conducted, based on the analysis of the scores obtained by nursing students at the University of Magdalena in the SABER PRO tests between 2016 and 2019. Descriptive and inferential analysis methods were used with Infostat/L software to evaluate the relationship between personal, family, and institutional variables and performance in the different modules of the test. It was found that better test performance is associated with personal variables such as younger age, male gender, and rural residence. Among the influential family variables are access to cable television, ownership of a computer and video game console, not residing in a permanent home, and not having dependents. At the institutional level, it was observed that the amount invested in tuition and scholarship holdings also affect the results.

The studies reviewed agree that student performance on the Saber Pro tests is influenced by multiple factors, including sociodemographic, academic, institutional, and personal factors. However, variables such as family environment, parents' educational level, socioeconomic status, gender, and mode of study are significant determinants of the results obtained. Research such as that by Osma Castellanos et al. (2014) and Timarán-Pereira et al. (2020) highlights the importance of socioeconomic and academic factors, while other studies such as that by Sánchez and Dahiana (2023) reinforce the influence of family context and residential environment. Likewise, statistical models and data mining techniques, such as those used by Narváez Zúñiga (2022) and Castillo Hernández (2013), have made it possible to identify performance patterns that reveal a marked relationship between personal conditions and academic results.

On the other hand, specific research by academic program, such as that of Acero (2017) and Canova-Barrios et al. (2023), shows that, in addition to general factors, there are particular dynamics depending on the discipline, region, and institutional profile. These studies highlight, for example, that variables such as age, parents' occupation, access to technological resources, or investment in tuition can also affect performance. Complementarily, studies such as those by Vásquez Arrieta (2018) and Rincón-Báez & Arias (2019) explore the relationship between previous tests, gender, and regional gaps, providing evidence on how the country's structural inequalities are reflected in the results of these standardized tests. Taken together, the evidence shows that performance on Saber Pro cannot be understood in isolation, but rather as a

multifactorial phenomenon that responds to the interaction of individual, social, and institutional conditions.

## 3. Methodology

### 3.1. Type of Research

A quantitative, descriptive, correlational, non-experimental research design was proposed. Statistical data from the Saber Pro tests were used and cross-referenced with demographic and socioeconomic information on students to identify patterns and significant relationships. The proposed methodology allowed for a structured and rigorous analysis of the behavior of Saber Pro test results in Colombia.

### 3.2. Population and Sample

Students from all programs nationwide took the Saber Pro tests in the years 2023-1 and 2024-1. The database contains information on 139,288 students.

### 3.3. Variables

| Dependent variable | Classification | Definition | Theoretical justification | Expected relationship |
|---|---|---|---|---|
| Overall score | Continuous quantitative | Numerical value representing the student's overall performance on the Saber Pro test, calculated based on the weighted average of competencies. | Dependent variable of the study. It is based on Human Capital Theory (Becker, 1964), which states that the competencies and knowledge acquired increase the value of the individual in the market and their academic performance. | This score is expected to be influenced by multiple factors such as grade point average, gender, modality, among others. |
| **Independent variables** | **Classification** | **Definition** | **Justification** | **Expected relationship** |
| Gender | Dichotomous categorical | Takes the value of 1 if the student is male. Zero otherwise | This variable is presented based on documents such as those by Vegas and Petrow (2007), Montalvo (2011), and Sun, Bradley, and Akers (2012). | According to the aforementioned studies, males perform better in mathematics and therefore could obtain a better overall score |
| Stratum | Ordinal categorical (1 to 6) | This is a proxy for income, with six categories. Stratum 1 is the lowest income and so on up to stratum 6, which is the highest income. Stratum 1 is the base variable. | This variable is presented based on documents such as Hanushek and Luque (2003), Escobar and Perdomo (2013), and Rodríguez, Ariza, and Ramos (2013). | Students from higher strata are expected to obtain better scores on Saber Pro. |
| Modality | Dichotomous categorical | Type of education: face-to-face (0) or distance learning (1). | This allows us to assess whether the educational modality has an effect on academic performance. | Students in face-to-face education are expected to perform better on average. |

| Mother's education | Polytomous categorical | Mother's level of education | This variable is presented based on the document by Narváez (2022) | It is expected that the higher the mother's level of education, the higher the test score will be. |
|---|---|---|---|---|
| Father's education | Polytomous categorical variable | Father's education level | This variable is presented based on the document by Narváez (2022). | It is expected that the higher the father's level of education, the higher the test score will be |
| Access to technological resources, specifically the internet | Dichotomous qualitative | Availability of technological resources for study | This variable is presented based on the document by Narváez (2022) | If these elements are available, a higher score will be obtained |
| Preparation method | nominal categorical | Three categories are presented that explain the preparation method: I do not use any preparation method (takes the value of 1), I review on my own (takes the value of 2), and I take a preparation course (takes the value of 3). I do not use any preparation method is the base variable. | This variable is presented to analyze whether ex ante test preparation methods are useful. No document in the literature was found that included it. | It would be expected that students who took a preparation course would have better results than those who did not use any method or reviewed on their own. The base variable is that I do not use any preparation method. |

## 3.4. Sources of information

The database of Saber Pro test results for all programs is obtained from the website:
https://www.icfes.gov.co/investigaciones/data-icfes/

## 3.5. Method of analysis

The **Ordinary Least Squares (OLS)** methodology is a fundamental technique in econometrics, used to estimate the parameters of a linear regression model. Its objective is to minimize the sum of the squares of the residuals, that is, the differences between the observed values of the dependent variable and the values predicted by the model. Its basic form is the simple linear regression model, which is expressed as:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon$$

where $y_i$ is the dependent variable, $x_i$ is the explanatory variable, $\beta_0$ and $\beta_1$ are the parameters to be estimated, and $\epsilon$ is the error term.
The OLS estimator seeks values $\beta_0$ and $\beta_1$ that minimize:

$$\sum (Y_i - \beta_0 - \beta_1 X_i)^2$$

Under the classical assumptions of the linear model (linearity, independence, homoscedasticity, no perfect multicollinearity, and zero-mean error), the OLS estimator is unbiased, consistent, and

efficient in the class of unbiased linear estimators, according to the **Gauss-Markov theorem** (Wooldridge, 2020).

Furthermore, when errors are normally distributed, OLS estimators are normally distributed and can be used to make statistical inferences (Gujarati & Porter, 2010). In practice, the design matrix and the use of matrix algebra allow the method to be extended to multivariate models, in which multiple coefficients are estimated simultaneously.

## 4. Results
## 4.1. Descriptive statistics

### Gender

| SEXO | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 78,861 | 56.62 | 56.62 |
| 1 | 60,427 | 43.38 | 100.00 |
| Total | 139,288 | 100.00 | |

### Stratum

| ESTRATO | Freq. | Percent | Cum. |
|---|---|---|---|
| 1 | 27,289 | 20.48 | 20.48 |
| 2 | 48,985 | 36.76 | 57.24 |
| 3 | 40,203 | 30.17 | 87.41 |
| 4 | 11,244 | 8.44 | 95.85 |
| 5 | 3,693 | 2.77 | 98.62 |
| 6 | 1,833 | 1.38 | 100.00 |
| Total | 133,247 | 100.00 | |

### Mother's education

```
. tab edu_madre_limpia
```

| edu_madre_limpia | Freq. | Percent | Cum. |
|---|---|---|---|
| | 2,914 | 2.09 | 2.09 |
| Educación profesional completa | 17,113 | 12.29 | 14.38 |
| Educación profesional incompleta | 4,303 | 3.09 | 17.47 |
| Ninguno | 4,316 | 3.10 | 20.57 |
| No Aplica | 595 | 0.43 | 20.99 |
| No sabe | 780 | 0.56 | 21.55 |
| Postgrado | 8,661 | 6.22 | 27.77 |
| Primaria completa | 10,582 | 7.60 | 35.37 |
| Primaria incompleta | 20,645 | 14.82 | 50.19 |
| Secundaria (Bachillerato) completa | 30,165 | 21.66 | 71.85 |
| Secundaria (Bachillerato) incompleta | 14,670 | 10.53 | 82.38 |
| Técnica o tecnológica completa | 18,071 | 12.97 | 95.35 |
| Técnica o tecnológica incompleta | 6,473 | 4.65 | 100.00 |
| Total | 139,288 | 100.00 | |

### Transformed data

| educ_madre_ordinal | Freq. | Percent | Cum. |
|---|---|---|---|
| Ninguno | 4,316 | 4.07 | 4.07 |
| Primaria | 31,227 | 29.42 | 33.48 |
| Secundaria | 44,835 | 42.24 | 75.72 |
| Profesional completa | 17,113 | 16.12 | 91.84 |
| Posgrado | 8,661 | 8.16 | 100.00 |
| Total | 106,152 | 100.00 | |

## Father's education

```
. tab edu_padre_limpia
```

| edu_padre_limpia | Freq. | Percent | Cum. |
|---|---|---|---|
| | 2,911 | 2.09 | 2.09 |
| Educación profesional completa | 16,075 | 11.54 | 13.63 |
| Educación profesional incompleta | 4,844 | 3.48 | 17.11 |
| Ninguno | 6,521 | 4.68 | 21.79 |
| No Aplica | 2,358 | 1.69 | 23.48 |
| No sabe | 3,543 | 2.54 | 26.03 |
| Postgrado | 7,923 | 5.69 | 31.71 |
| Primaria completa | 10,567 | 7.59 | 39.30 |
| Primaria incompleta | 25,724 | 18.47 | 57.77 |
| Secundaria (Bachillerato) completa | 27,410 | 19.68 | 77.45 |
| Secundaria (Bachillerato) incompleta | 13,422 | 9.64 | 87.08 |
| Técnica o tecnológica completa | 12,894 | 9.26 | 96.34 |
| Técnica o tecnológica incompleta | 5,096 | 3.66 | 100.00 |
| Total | 139,288 | 100.00 | |

| educ_padre_cat5 | Freq. | Percent | Cum. |
|---|---|---|---|
| Ninguna | 6,521 | 4.68 | 4.68 |
| Primaria | 36,291 | 26.05 | 30.74 |
| Secundaria | 40,832 | 29.31 | 60.05 |
| Profesional | 38,909 | 27.93 | 87.99 |
| Posgrado | 7,923 | 5.69 | 93.67 |
| . | 8,812 | 6.33 | 100.00 |
| Total | 139,288 | 100.00 | |

## Has internet

| tiene_internet | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 15,777 | 11.57 | 11.57 |
| 1 | 120,640 | 88.43 | 100.00 |
| Total | 136,417 | 100.00 | |

## VIRTUAL  IN-PERSON

| metodo_num | Freq. | Percent | Cum. |
|---|---|---|---|
| 0 | 3,672 | 24.33 | 24.33 |
| 1 | 11,422 | 75.67 | 100.00 |
| Total | 15,094 | 100.00 | |

### 4.2. Regression model.

A regression is proposed that examines how gender, socioeconomic status, mother's and father's education (as sociodemographic conditions), internet availability, preparation method, whether it is a face-to-face or distance learning program, and preparation method (as institutional conditions) influence the results of the Saber Pro tests.

$$PG= \beta_0 + \beta_1 \text{ GENDER} + \beta_2\text{STRATA} + \beta_3\text{MODALITY} + \beta_4\text{MOTHER'S EDUCATION} + \beta_5\text{FATHER'S EDUCATION} + \beta_6\text{INTERNET} + \beta_7\text{PREPARATION METHOD} + \text{€i}$$

The results of the model are shown below

```
. reg PUNTAJE_GLOBAL i.SEXO i.ESTRATO  i.EDU_MADRE_FINAL i.EDU_PADRE_FINAL i.INTERNET i.VIRTUAL_PR      (1)
> ESENCIAL i.COMO_SE_CAPACITO, robust
```

Linear regression

| | | | | | Number of obs | = | 10,544 |
|---|---|---|---|---|---|---|---|
| | | | | | F(18, 10525) | = | 133.35 |
| | | | | | Prob > F | = | 0.0000 |
| | | | | | R-squared | = | 0.1843 |
| | | | | | Root MSE | = | 21.218 |

| PUNTAJE_GLOBAL | Coefficient | Robust std. err. | t | P>\|t\| | [95% conf. interval] | |
|---|---|---|---|---|---|---|
| 1.SEXO | 1.575944 | .4263436 | 3.70 | 0.000 | .7402299 | 2.411658 |
| **ESTRATO** | | | | | | |
| 2 | 3.587765 | .5474013 | 6.55 | 0.000 | 2.514755 | 4.660775 |
| 3 | 5.618513 | .60959 | 9.22 | 0.000 | 4.423601 | 6.813424 |
| 4 | 11.59938 | .9853757 | 11.77 | 0.000 | 9.667859 | 13.53091 |
| 5 | 12.73892 | 1.659024 | 7.68 | 0.000 | 9.486923 | 15.99093 |
| 6 | 14.32038 | 2.135245 | 6.71 | 0.000 | 10.1349 | 18.50587 |
| **EDU_MADRE_FINAL** | | | | | | |
| Primaria | 4.429968 | 1.247391 | 3.55 | 0.000 | 1.984845 | 6.875091 |
| Secundaria | 8.31414 | 1.262963 | 6.58 | 0.000 | 5.838494 | 10.78979 |
| Profesional completa | 12.84296 | 1.404982 | 9.14 | 0.000 | 10.08893 | 15.59699 |
| Posgrado | 16.97428 | 1.578753 | 10.75 | 0.000 | 13.87962 | 20.06893 |
| **EDU_PADRE_FINAL** | | | | | | |
| Primaria | .8765134 | 1.024101 | 0.86 | 0.392 | -1.130919 | 2.883946 |
| Secundaria | 1.142494 | 1.056491 | 1.08 | 0.280 | -.9284293 | 3.213418 |
| Profesional | 3.254974 | 1.132263 | 2.87 | 0.004 | 1.035525 | 5.474424 |
| Posgrado | 9.926733 | 1.472506 | 6.74 | 0.000 | 7.040342 | 12.81312 |
| 1.INTERNET | 3.367652 | .6456725 | 5.22 | 0.000 | 2.102012 | 4.633292 |
| 1.VIRTUAL_PRESENCIAL | 7.9908 | .4614144 | 17.32 | 0.000 | 7.08634 | 8.89526 |
| **COMO_SE_CAPACITO** | | | | | | |
| Repasó solo | -3.555991 | .5250681 | -6.77 | 0.000 | -4.585224 | -2.526758 |
| Tomó curso | -5.26821 | .9890369 | -5.33 | 0.000 | -7.206909 | -3.32951 |
| _cons | 121.9994 | 1.22946 | 99.23 | 0.000 | 119.5894 | 124.4093 |

## 5. Discussion of the results obtained.

The results of the model are described below

**Gender:**

The gender variable was significant. The gender of students does influence the results by between one and two points. Various studies have reported similar findings . For example, Valverde and Naslund-Hadley (2010) found these types of differences for El Salvador and Colombia. Valens (2007) found the same result for economics programs in the former ECAES, and Pineros (1988) noted this at the secondary level.

**Stratum:**

At the socioeconomic level, the stratum variable was significant for all strata.

According to the regression, each stratum will score the following additional points on average on the Saber Pro tests when compared to stratum one:

People in stratum two will have an average of three more points on the Saber Pro tests than those in stratum one.

People in stratum three will have an average of five more points on the Saber Pro tests than those in stratum one.

People in stratum four will score an average of eleven points higher on the Saber Pro tests than those in stratum one.

People in stratum five will score an average of twelve points higher on the Saber Pro tests than those in stratum one.

People in stratum six will score an average of fourteen points higher on the Saber Pro tests than those in stratum one

**Mother's education**

The results indicate that if the mother has a primary education, she will score 4 points more; if she has a secondary education, 8 points more; if she is a professional, 12 points more; and if she has a postgraduate degree, 16 points more, when compared to students whose mothers have no education. All coefficients were significant

**Father's education**

Primary and secondary education are not significant. If the father is a professional, the student will score an average of 3 points more, and if he has a postgraduate degree, 9 points more.

**Study mode**

The variable referring to the mode of study (virtual and face-to-face) was significant, and it was found that face-to-face programs score an average of 7 points more when compared to virtual programs.

**Internet availability**

Internet availability was significant and led to an average of three more points on the Saber Pro tests compared to those who did not have this technological resource.

**Training method**

The categories "trained alone" and "took courses" have negative values. In other words, those who studied on their own or took courses obtained lower scores than those who did not prepare, which defies all logic.

A summary of the results is presented in the table

| Variable | Result | Interpretation |
|---|---|---|
| Gender | Significant (+1 to 2 points for men) | There is a slight performance advantage for men. |

| Stratum | All significant (up to +14 points in stratum 6) | Higher stratum, higher academic performance. |
|---------|------------------------------------------------|----------------------------------------------|
| Mother's education | Significant at all levels (up to +16 points) | Higher maternal education improves results. |
| Father's education | Only professional and postgraduate degrees are significant | Weaker effect than that of the mother. |
| Study mode | Face-to-face +7 points | Face-to-face learning is associated with better performance. |
| Internet access | Significant (+3 points) | Internet access improves scores. |

## 6. Discussion.

The model results confirm that socioeconomic and institutional conditions continue to play a decisive role in students' academic performance on the Saber Pro tests. The evidence shows that gender differences, although slight, persist over time and are expressed in advantages for men, as pointed out by Valverde and Naslund-Hadley (2010) and Valens (2007). Likewise, socioeconomic status continues to be a decisive factor, as students from higher levels obtain better scores, reflecting the persistence of structural inequalities in the Colombian education system.

On the other hand, the results confirm that the mother's educational level has a greater than the father's on academic achievement, highlighting the role of the family influenceenvironment in the student's education. It is also noteworthy that face-to-face learning and internet access are associated with better performance, highlighting the importance of strengthening the technological and pedagogical infrastructure of institutions. Finally, the findings on preparation methods highlight the need to review the quality and relevance of the courses offered, so that they contribute effectively to the development of skills and are not limited to superficial training.

## 7. Conclusions

The research found that socioeconomic and institutional conditions do indeed influence the results of the Saber Pro tests at the national level, and a series of policies can be suggested to improve performance.

Firstly, there is a slight gender gap in favor of men. The literature reviewed finds that these gaps are present from primary and secondary school levels, therefore, a public policy that emphasizes closing this gap should be considered.

On the other hand, the study confirms that students from higher socioeconomic levels obtain better results. This confirms that it is still necessary to invest in improving the quality of education, especially public education, so that schools themselves can improve the skills of students from lower socioeconomic strata and this is not reflected later on in university. The variable of internet access reinforces these results, as the mere availability of this element increases the average results obtained.

Some public policies that can be suggested are:
-Implementing strategies to close the gender gap starting in basic education.
- Investing in educational quality in lower socioeconomic strata, especially in public institutions.
- Design educational training programs for parents with low levels of schooling.
- Strengthen face-to-face learning and improve the quality of virtual education.
- Guarantee universal access to the internet and technological resources.
- Offer preparation courses for the Saber Pro test with public subsidies.

Finally, future lines of research are suggested to further explore these results.

☐ Longitudinal studies from Saber 11 to Saber Pro.

☐ Detailed evaluation of preparation methods.

☐ Analysis of interactions between variables (e.g., gender and socioeconomic status).

☐ Inclusion of institutional variables (type of university, accreditation).

☐ Regional analysis and use of techniques such as Random Forest or quantile regression.

## References.

Acero Báez, Y. D. (2017). Factores demográficos, socioeconómicos y académicos que influyen en los resultados del componente genérico de la Prueba Saber Pro-Caso: Ingeniería Electromecánica de la Universidad Pedagógica y Tecnológica de Colombia Facultad Seccional Duitama.

Bahamón Muñetón, M. J., & Reyes Ruiz, L. (2014). Caracterización de la capacidad intelectual, factores sociodemográficos y académicos de estudiantes con alto y bajo desempeño en los exámenes Saber Pro—Año 2012. Avances en Psicología Latinoamericana, 32(3), 8.

Canova-Barrios, Carlos Jesús; Acosta-Salazar, Diana Patricia; Torres-Barrios, Leonardo Andrés y Alvarez-mino, Lídice (2023). Variable que influyen sobre el desempeño de los estudiantes de enfermería de la Universidad del Magdalena en las pruebas SABER PRO. 2023, vol.12, n.2.

Cifuentes, J. (2013) Predicción del resultado en la prueba SABER PRO para Economía a partir de la información disponible en el proceso de admisión. Trabajo de grado para optar por al título de magister en Economia. Universidad Javeriana.

Demarchi Sánchez, G. D. (2023). Factores que intervienen en el desempeño de los estudiantes en las Pruebas de Estado Saber Pro. El Ágora USB.

Escobar Pineda, S., & Orduz Perdomo, M. (2013). Determinantes de la calidad en la educación superior en Colombia. Tesis de maestría en Economía. Universidad javeriana.

Hanushek, E. A., & Luque, J. A. (2003). Efficiency and equity in schools around the world. Economics of Education Review, 22(5), 481–502.

Gujarati, D. N., & Porter, D. C. (2010). *Econometría* (5.ª ed.). McGraw-Hill.

Melo-Becerra, Ligia Alba; Ramos-Forero, Jorge Enrique; Hernández-Santamaría, Pedro Oswaldo. "La educación superior en Colombia: situación actual y análisis de eficiencia". Revista Desarrollo y Sociedad. N.º 78, 2017, pp. 59-111.

Narvaez-Zuniga (2023) "Modelo estadístico para determinar los valores académicos en los resultados de las pruebas saber pro". Revista investigación e innovación en ingenierías. Vol. 11. No 1, 3-21.

Ramírez, Carla. (2014). Factores asociados al desempeño académico según nivel de formación pregrado y género de los estudiantes de educación superior Colombia. Revista Colombiana de Educación, (66), 203-224.

Rincón-Baez, W. U., & Arias-Velandia, N. (2019). Brecha de rendimiento académico por género en saber pro en programas de administración en los departamentos de colombia academic achievement gender gap in saber pro through administration programs in the departments of colombia. Panorama, 13(2 (25)), 142-161. https://doi.org/10.15765/pnrm.v13i25.1410

Rodríguez, G., & Ariza, M. (2014). Determinantes del desempeño académico universitario. El caso de la Región Caribe colombiana. ICFES. Saber Pro. Bogotá, DC.

Sun, L., Bradley, K. D., & Akers, K. (2012). A Multilevel Modelling Approach to Investigating Factors Impacting Science Achievement for Secondary School Students: PISA Hong Kong Sample. International Journal of Science Education, 1-19.

Timarán Pereira, R., Hidalgo Troya, A., & Vidal Alegría, F. (2020). Una mirada al desempeño académico en las pruebas saber pro de los estudiantes de ingeniería desde la minería de datos educativa. RISTI: Revista Ibérica de Sistemas e Tecnologias de Informação, Extra 38, 29-42. https://dialnet.unirioja.es/servlet/articulo?codigo=8574384

Pineros. L y Rodrigues, A. (1988). Los insumos escolares en la educación secundaria y su efecto sobre el rendimiento académico de los estudiantes. Un estudio en Colombia. Banco Mundial, departamento de desarrollo humano. Paper series no 36. Diciembre.

Valens (2007). Calidad de la educación superior en Colombia: un análisis multinivel con base en el ECAES de economía de 2004. Revista Sociedad y Economia. Universidad del valle.

Valverde y Naslund-Hadley (2010). La condición de la educación en matemáticas y ciencias naturales en America latina y el caribe. División de educación. Notas técnicas. Banco interamericano de desarrollo. Washington.

Vásquez Arrieta, O. (2018). Las pruebas SABER 11° como predictor del rendimiento académico expresado en los resultados de la prueba SABER PRO obtenidos por las estudiantes de la Licenciatura en Pedagogía Infantil de la Corporación Universitaria Rafael Núñez. Hexágono Pedagógico, 9(1), 187–204.

Vegas, E. & Petrow, J., 2007. Raising student achievement in Latin America: the challenge for the 21st century. Washington, D.C.: World Bank Publications.

Wooldridge, J. M. (2020). *Introductory econometrics: A modern approach* (7th ed.). Cengage Learning.