Gómez Soler, Silvia C.; Bernal Nisperuza, Gloria L; Herrera Idárraga, Paula
TEST PREPARATION AND STUDENTS' PERFORMANCE:
THE CASE OF THE COLOMBIAN HIGH SCHOOL EXIT EXAM*

TEST PREPARATION AND STUDENTS' PERFORMANCE: THE CASE OF THE COLOMBIAN HIGH SCHOOL EXIT EXAM*

Silvia C. Gómez Soler
Gloria L. Bernal Nisperuza
Paula Herrera Idárraga

S. C. Gómez Soler

PhD in Policy Analysis. Assistant Professor, Pontificia Universidad Javeriana, Bogotá, Colombia. e-mail: silvia.gomez@javeriana.edu.co

G. L. Bernal Nisperuza

MPA. Assistant Professor, Pontificia Universidad Javeriana, Bogotá, Colombia. e-mail: gbernal@javeriana.edu.co

P. Herrera Idárraga

PhD in Economics. Associate Professor, Pontificia Universidad Javeriana, Bogotá, Colombia. e-mail: pherrera@javeriana.edu.co

Gómez Soler, S. C., Bernal Nisperuza, G. L., & Herrera Idárraga, P. (2020). Test Preparation and Students' Performance: The Case of the Colombian High School Exit Exam. *Cuadernos de Economía*, **39(79), 31-72.**

Standardized test scores play a central role in determining college admission decisions in both developed and emerging countries. For that reason, many courses are offered by schools and other educational institutions to prepare students for the exams. However, it is still unclear whether additional preparation has a positive and significant effect. The objective of this paper is to use the results of the Colombian high-school exit examination to gain a better understanding of the relationship between preparation and test results. Our results show that, on average, preparatory activities are associated with an increase of approximately 0.06 standard deviations in scores.

Keywords: exit exams, economics of education, test preparation, shadow education.
JEL: I20, I21, I24, I25, I26.

Gómez Soler, S. C., Bernal Nisperuza, G. L., & Herrera Idárraga, P. (2020) La preparación y el desempeño: el caso de las pruebas de Estado en Colombia. *Cuadernos de Economía*, **39(79), 31-72.**

Los resultados de los exámenes estandarizados juegan un papel importante en las decisiones de admisión a la universidad en países desarrollados, emergentes y de bajos ingresos. Por esa razón, los colegios y otras instituciones educativas ofrecen cursos para preparar a los estudiantes. Sin embargo, no está claro si esta preparación adicional tiene un efecto positivo y significativo. Este artículo utiliza los resultados del examen de Estado colombiano para analizar la relación entre la preparación y el desempeño. Nuestros resultados muestran que, en promedio, las actividades de preparación están asociadas a un incremento de aproximadamente 0.06 desviaciones estándar en los puntajes.

Palabras clave: examen de Estado, economía de la educación, preparación para el examen, educación sombra.
JEL: I20, I21, I24, I25, I26.

Gómez Soler, S. C., Bernal Nisperuza, G. L., & Herrera Idárraga, P. (2020) La préparation et les résultats: le cas des épreuves d'Etat en Colombie. *Cuadernos de Economía, 39(79), 31-72.*

Les résultats des examens standardisés jouent un rôle important dans les décisions d'admission à l'université dans les pays développés, émergeants et de faibles revenus. Pour cette raison, les établissements secondaires et autres institutions éducatives offrent des cours pour préparer les étudiants. Cependant, il n'est pas certain que cette préparation additionnelle ait un effet positif et significatif. Cet article utilise les résultats de l'examen d'Etat colombien pour analyser la relation entre la préparation et les résultats. Notre étude montre que, en moyenne, les activités de

préparation sont associées à une augmentation d'environ 0,06 des résultats standards dans les notes.

Mots-clefs: examen d'Etat, économie de l'éducation, préparation pour l'examen, éducation de l'ombre.
JEL: I20, I21, I24, I25, I26.

Os resultados dos exames padronizados desempenham um papel importante nas decisões de admissão à universidade em países desenvolvidos, emergentes e de baixos ingressos. Por essa razão, os colégios e outras instituições educativas oferecem cursos para preparar aos estudantes. No entanto, não está claro se esta preparação adicional tem um efeito positivo e significativo. Este artigo utiliza os resultados do exame de Estado colombiano para analisar a relação entre a preparação e o desempenho. Nossos resultados mostram que, em média, as atividades de preparação estão associadas a um incremento de aproximadamente 0.06 desvios padrão nas pontuações.

Palavras-chave: exame de Estado, economia da educação, preparação para o exame, educação sombra.
JEL: I20, I21, I24, I25, I26.

INTRODUCTION

Standardized test scores play a central role in determining college admission decisions in developed and emerging countries. As Chetty et al. (2011) have shown, the results obtained in these tests are potentially correlated with future academic and labor opportunities. For that reason, there has been a substantial increase in the courses offered by schools and other educational institutions to prepare students for these exams.

In recent years, supplementary tutoring, also known as shadow education[1], has become a common practice in the United States of America and several countries in Asia (China, Republic of Korea, South Korea, Taiwan, and Vietnam) and Europe (Turkey and Switzerland). This practice has caught the attention of educational researchers, as it is unclear whether additional preparation has a positive and significant effect on academic achievement, even though it is a common, time-consuming and often expensive practice.

It has also been of interest to policymakers because only some students have access to preparatory activities, potentially creating a gap between students from different socioeconomic backgrounds. If this is the case, the preparatory activities could be exacerbating the already unequal access to higher education for students who are at risk. The massification of supplementary tutoring has also been a concern for schools, parents and government officials, since these activities may generate perverse incentives for schools to teach to the test, disregarding other essential factors of education, and for teachers to provide lower-quality education within the school system, leaving the responsibility of teaching to coaches outside school (Jayachandran, 2014).

In an attempt to accurately estimate the effect, there is a large body of literature that studies the effect of test preparation on student performance using an ample variety of methodologies (instrumental variable, propensity score matching and nonparametric methods, among others). However, the results of these studies have not been conclusive. Besides, few studies have attempted to quantify the effect of getting prepared at different types of institutions, and whether the duration, it is an important factor in assessing the effectiveness. Furthermore, the literature that studies the heterogeneous effects of preparation by socioeconomic strata and type of school (public vs private) is very scarce.

To the best of our knowledge, no efforts have been made to econometrically quantify the effect of test preparation on achievement for a Latin American country. The objective of this paper is to gain a better understanding of the relationship between test preparation and the results in the Colombian high school exit examination (Saber11). This paper also aims to assess the heterogeneous effects of preparation on specific subgroups (gender, socioeconomic strata, type of school). Additionally, this study contributes to the literature by examining the effect of get-

---

[1] see Bray (2013) for an extensive overview of this phenomenon.

ting prepared at different sites (schools, test preparation institutions, universities) and analyses whether the duration of these activities may lead to different results.

The rest of this paper is organized as follows: Section 2 reviews the academic literature that has studied the effects of test preparation on achievement. Section 3 describes the Colombian educational system. Section 4 presents an overview of preparatory activities for the Saber11 examination. Section 5 describes the theoretical framework used in this article. Section 6 presents descriptive statistics of the dataset used for the analysis. Section 7 describes the methodology employed for the analysis and discusses the specification of the model. Section 8 reports the main results. Section 9 presents the results of the robustness checks. Section 10 provides a discussion of the main results and policy implications, and Section 11 presents conclusions.

LITERATURE REVIEW

In the United States, a country with a long tradition administering standardized tests, there is a consensus that participation in preparatory activities can lead to an increase in SAT scores. However, the magnitude of the effect is still a matter of debate. According to Briggs (2002), at least thirty-four studies have been published since 1953. evaluating the effect of coaching on SAT scores. For instance, Domingue and Briggs (2009) use a representative longitudinal sample of high school students in the United States and propensity score matching to determine the effect of preparatory activities on SAT scores. The results of Domingue and Briggs (2009) indicate that students who took the PSAT (equivalent to a mock SAT) and had some other type of preparation scored higher in math. However, they did not find a statistically significant effect on language. Buchmann, Condron, & Roscigno, (2010) find that students from the most-advantaged families are more likely to engage in private tutoring, and the tutoring they receive is of the highest quality. Byun and Park (2012) assess the effect of two types of private tutoring: commercial test preparation courses and one-to-one tutoring, on SAT performance for different racial/ethnic students. Using OLS regressions, the authors find that commercial SAT preparation courses have a positive and significant effect only for East Asian American students, and one-to-one tutoring did not display any effect for any of the racial/ethnic students.

Several studies have also examined the effect of private tutoring on academic performance outside of the United States. Dang (2007) uses household survey data, containing information about household expenditures in private tutoring and (self-reported) academic performance of students in Vietnam, to analyze the effects of preparatory activities on exam performance. Dang (2007) finds that private tutoring can help improve the performance of primary and secondary students. For the case of South Korea, using OECD's PISA results, Choi, Calero, & Escardíbul. (2011) find that the effect of the time spent in private tutoring activities depends on the specific subject that is being examined. The results show that the effect is

positive for mathematics, positive but decreasing for reading, and non-significant for science. In a similar study, Ryu and Kang (2013) also examine the causal effect of private tutoring expenditures on the academic performance of middle school students in South Korea using instrumental variables. The results indicate that higher expenditures on private tutoring have a positive but fairly small effect. Ryu and Kang's (2013) results reveal that a 10-percent increase in overall expenditure is related to an improvement of 0.03 SD (Ryu & Kang, 2013).

For China, Zhang (2013) uses an instrumental variable approach and a hierarchical linear model to show that private tutoring has a mixed and heterogeneous effect on mathematics and language in the National College Entrance Exam (NCEE). Zhang and Liu (2016) also study the Chinese case using a propensity score matching approach, differentiating between four types of private tutoring according to the size of the tutoring class. Their results reveal that, in general, private tutoring has no significant effects on the results obtained in the NCEE, even after accounting for class size. Only large class size tutoring in mathematics has a significantly positive effect on the NCEE scores, and only for rural students is the effect quite considerable (0.248 S.D.).

For the case of students in junior high schools in Taiwan, Kuan (2011) observes a small effect that mathematics cramming programs have on mathematics performance. This researcher also shows that students who are more likely to attend cram schools benefit less from math cramming compared to students who are less likely to attend or who come from less-advantageous social backgrounds. Liu (2012) also finds that cram schooling has a positive effect on analytical ability and math performance. However, the positive effects of cram schooling decrease when cram schooling hours are extended.

Studying the effect of private tutoring for a European country, Hof (2014) finds a heterogeneous and nonlinear effect on the academic outcomes for Swiss students in reading and mathematics in the PISA-2009 examination using nonparametric methods. Accounting for non-random selection, Gurun and Millimet (2008) assess the causal effect of private tutoring on the probability of university placement in Turkey. Their results show that private tutoring has a negative causal effect on a university placement.

As reviewed in this section, several researchers have contributed to the literature that studies the relationship between test preparation activities and levels of achievement. However, there is still much ground to cover in this area of research. As was highlighted throughout this review of the literature, there are mixed results in the existing literature, even within a country, that justifies more research on this topic. Furthermore, preparation activities are very different in each of the cases studied in the literature. The activities conducted in tutoring sessions differ in type, quality and quantity (for a more comprehensive discussion, see Bray, 2014).

Additionally, the evidence for Latin America is very scarce. To the best of our knowledge, only one effort has been made to try to understand the effect of preparation

on achievement for the Colombian case. Such an effort was made by Hernandez (2011). This researcher uses the scores from the 2009 Saber11 examination and the results from a survey conducted in schools to demonstrate that there is a relationship between test performance and the type of preparation. The results reveal that performance depends on the degree of involvement in the test preparation activity offered at the school, and not on the socio-cultural context of the individuals enrolled at the educational institution. However, these results cannot be used to determine the effect of preparation on academic performance quantitatively.

Therefore, this paper contributes to the literature in several ways. First, we contribute to the search for new and better ways to understand the relationship between test preparation and educational achievement in a Latin American context. Second, we quantify the effect of preparation in academic performance by characterizing the effect of the different institutions that offer preparation activities and by incorporating into the analysis the duration of the preparatory activities. Third, unlike previous studies, we use census data of students who took the same exam on the same day, which allows us to obtain reliable and generalizable estimates for the population studied.

COLOMBIAN CONTEXT

The Colombian Educational System is divided into four levels: pre-school, basic (primary and secondary), upper secondary and higher education. The first compulsory level of education in Colombia is pre-school. Students usually enroll in pre-school at age 5. At age 6. students go on to basic education (Grades 1-9 for 6-14-year-olds) where they complete two sub-levels: primary school (*Educación Primaria* - 5 academic years) and lower secondary school (*Educación Secundaria* - 4 academic years). At age 15-16. students start upper secondary school (*Educación Media* – 2 academic years), and upon successful completion, they receive the *bachiller* certificate. In our dataset, the average age of students at graduation is 16.83 years old (*Table 1*).

Some students enrolled in upper secondary school attend classes in the morning and the afternoon (full time - *jornada única*). However, other students go to school only for half a day, either in the morning or the afternoon, at night or during the weekends. The students enrolled in a full-time schedule spend on average 8 hours at school every day, whereas students enrolled in a morning, afternoon or night shift only spend 4 to 5 hours per day.

To qualify for higher education studies, all students must take the Saber11 exam. According to Colombian Law[2], High School Principals have to report a list of all students who are currently enrolled in the last year of Upper Secondary School so that they get registered to take the Saber11 examination. The Colombian nation-wide

---

[2] Article 5th of Act 869 of 2010.

high school exit exam, administered by the Colombian Institute for Educational Evaluation (ICFES), was first established more than forty years ago in order to provide support to universities and other higher education institutions in their admissions processes. Throughout these years, the exam has gone through a series of significant changes to capture the needs of the educational system in Colombia. The exit examination was initially conceived as an exam to test specific contents, but in the year 2000, it became an exam to test competencies.

The Saber11 exam is divided into two main components: the core subject component (eight subjects) and the flexible component. The core subject's component is compulsory for all test takers, and it evaluates language (Spanish), mathematics, biology, chemistry, physics, social sciences, philosophy, and English. The flexible component has two different types of tests: advanced skills and interdisciplinary. The advanced skills exam tests one of the core subjects at a higher level of complexity. Students can choose from language (Spanish), mathematics, biology and history. The flexible component exam tests multidisciplinary problem-solving skills. The results for each of the core subjects are reported on a 0 to 100 scale, with an average score of 50 and a standard deviation of 10 (ICFES 2010). A global score that measures the performance in all areas of the exam is also calculated as the weighted average of the score obtained in all the subjects[3]. According to ICFES guidelines, the scores are calculated using a standardization process constructed using Item Response Theory (IRT).

Tertiary institutions are autonomous to decide their admissions criteria. However, the most selective universities in Colombia use the Saber11 results as their only criteria. Other universities use a set of admissions criteria to determine admittance to a program, including a combination of Saber11 results, interviews and other specific tests.

PREPARATORY ACTIVITIES FOR THE SABER11 EXAMINATION

Earning a high score on the Saber11 examination is crucial because these results are used in Colombia as a filter to be admitted to higher levels of education, to qualify for scholarships and to be accepted when applying for jobs. Every year, the Ministry of Education organizes a special event, The Night of the Best (*La Noche de los Mejores*), to recognize those students who obtained the highest scores in the Saber 11 examination. Fifty students receive a full scholarship that includes tuition, board, books and transportation. High school officials are also interested in the Saber11 results because these scores are used to construct rankings and are often interpreted as a measure of the school's quality. For instance, the Ministry of Education constructs an Index, The Synthetic Index of Education Quality (*Índice*

---

[3] global score = (Mathematics*3+ Language*3+ Social Sciences*3+ Philosophy*3+ Biology*3+ Chemistry*3. Physics and English*1)/19

Sintético de Calidad Educativa – ISCE), to rank schools based on the results that students obtain in the Saber examinations (Ministerio de Educacion 2017). The Colombian media have also used the results of the Saber11 exam to construct rankings of what they consider to be the best academic programs. A nationally recognized magazine, Revista Dinero (2017), publishes a special issue every year titled "The Best Schools in Colombia According to the Saber11 exam". This magazine ranks the quality of schools using the average results obtained by students at different institutions in the Saber11 exam. Individuals' perceptions of the education offered at different schools are influenced by these rankings, and the parents of prospective students often incorporate this information in their decision-making process. In sum, these rankings can affect the reputation of a given institution.

Students enroll in preparatory activities with the hope of achieving higher scores. The main preparatory activities offered to students are courses at i) a school with a schoolteacher, ii) a school with external faculty, iii) a university and iv) a private educational institution commonly known as a pre-ICFES institute. The courses offered by a university or a pre-ICFES institute are considerably more expensive and require more time and dedication from families and students in comparison to preparation activities at school. Some students also choose to take the Pre-Saber exam, which is a diagnostic test that students can take approximately six months before taking the Saber11 exam. Both tests are similar in structure and are administrated by ICFES.

A considerable percentage of students in Colombia get prepared at school. Schools have incentives to offer these courses because their quality is classified based on the Saber11 scores obtained by their students (rankings). Furthermore, private schools use these rankings for advertisement purposes if they do well. Public schools that perform well in the exam use the results to obtain additional resources from the government. The percentage of students who get prepared at their own schools is around 70%. The remaining 30% get enrolled in other courses that have a cost such as the courses offered at universities and pre-icfes institutes[4].

In the year 2012. ICFES decided to include a survey that students had to answer the day of the exam with questions about the preparatory activities. Those students who took the examination had to fill out the survey before the test started. The mandatory survey contained questions about the following: i) whether the student prepared for the test, ii) the institution where the preparatory course took place (at school with one of the school's teachers, at school with external faculty, at a University or a private educational institution commonly known as a pre-ICFES institute), and iii) the time spent on preparation. It should be acknowledged that this information comes from students' self-reports. It is possible that some students stated being prepared when in fact they did not, and those who received preparation might not have reported it. This reporting error seems more likely in the case

---

[4] Author's calculations using information from a survey about preparation activities administered by ICFES in 2012.

of students who were prepared at school by their teachers. Some students may not recognize the difference between a preparatory activity and their regular classes. If preparation has a positive effect on the result, then these errors in self-reported information would generate a negative bias. Unfortunately, with the available data, it is neither possible to detect nor correct this error. According to this survey, the activities of the different preparatory courses are relatively homogenous. The teachers usually provide explanations about a particular topic, and then students practice answering multiple choice questions. Most of the courses also include a mock exam. Nevertheless, the survey showed that the intensity of the courses in terms of months differed among types of preparation.

The costs of preparatory activities can be relatively high in Colombia. For instance, the cost of a preparatory course offered at a certified Pre-ICFES institute can be as high as 4´800,000 Colombian Pesos (using an exchange rate of 2.900 COP/USD, it is approximately 1,724 USD), which is approximately 7.8 times the monthly minimum wage in Colombia. Furthermore, the cost of preparatory courses at universities can range from 400,000 COP to 1.700,000 COP, which is still an exorbitant cost for a family that earns one or two minimum wages. According to a survey conducted by ICFES in 2014. the cost of preparation at schools varies widely depending on the school, and it is the less-expensive alternative. Students pay on average between 200,000 COP and 300,000 COP for courses offered at school by external faculty[5]. These costs can vary dramatically by location. Capital cities are more likely to offer specialized courses and, usually, have the most expensive programs.

A pre-test, known as Pre-Saber, could be taken by anyone who signed up and paid for it, and it has the same duration and structure of the Saber11 examination. The cost of the Pre-Saber examination is approximately 56,000 Colombian pesos (which is approximately the equivalent of two daily minimum wages in Colombia). Some students, especially those facing financial difficulties, are unable to pay for these preparatory activities, and therefore, they only receive preparation at their own schools or do not receive preparation at all.

THEORETICAL FRAMEWORK

The literature on the economics of education provides models of schooling decisions using an educational production function. We consider the education production function model, proposed by Hanushek (1979) and Todd and Wolpin (2003), as the framework for estimating the effect of test preparation activities on students' performance on the Colombian High School Exit Examination. Following Todd and Wolpin's (2003) notation, let T_i be a measure of achievement for student i. In this model, the inputs are given by the choices that are being made by parents and

---

[5] Based on a voluntary survey about preparation activities for Saber11 administered by ICFES to 38.955 students in 2014.

schools. F_i represents the vector of inputs supplied by the parents at a given age, and S_i represents the vector of school-supplied inputs. D_i denotes external factors. In this case, the preparatory activities affecting student i are modeled so that they are included as part of the inputs supplied by the parents and schools. The student's endowed ability is denoted as μ_i and a measurement error in test scores as ε_i. The production function is represented as:

$$T_i = T\left(F_i, S_i, D_i, \mu_{i0}, \varepsilon_i\right) \tag{1}$$

Missing data for past inputs can become problematic at the time of estimation. To overcome that problem, Todd and Wolpin (2003) formulated a specification that only depends on contemporaneous measures of school and family inputs. This contemporaneous specification is adopted in many empirical studies because of the limited availability of historical input measures or historical test score results. The implementation of a contemporaneous specification assumes that ε_i is an additive error that includes any omitted factors, including past inputs and endowed capacity, and the possibility of measurement error. Therefore, the residual term in equation (1) includes all the omitted factors.

To date, the literature has identified several factors that are considered to be determinants of students' educational outcomes. Hanushek (1986) points out that the inputs included in an educational production function should be relevant to the students under study. Following the structure of the educational production function described previously, in this research educational achievement of a given student at a particular point of time is presented as a function of the characteristics of the student's family, the characteristics of the schools and teachers, and innate abilities of the student (Hanushek 1986).

Family background variables can play a crucial role in educational achievement (Haveman & Wolfe 1995; Häkkinen, Kirjavainen, & Uusitalo, 2003; Woßmann, 2003). In this article, we included two variables to account for family background: parental education and socioeconomic strata. Parental education is incorporated into the analysis because Woßmann (2003) and Häkkinen et al. (2003) have shown that students whose parents have a higher level of education have better academic results. This could be explained by the genetic transmission of skills from parents to kids, and the cultural environment that highly educated parents provide, which fosters learning. The socioeconomic strata, our proxy for the income level of the families, is also essential for the analysis because as Davis-Kean (2005) and Belley and Lochner (2007) have shown, family income is also related to achievement.

According to Gamoran and Long (2007), the characteristics of the school also have an important impact on academic performance. In this article, several variables, for which there is data availability, are included to account for school characteristics. Some of the variables included are school location, school meeting time (shift) and school type (public/private).

The preparatory activities affecting student *i* are modeled so that they are included as part of the inputs supplied by the parents and schools affecting the educational production function. A large body of literature, including the contributions of Briggs (2002), Domingue and Briggs (2009), and Dang (2007), have studied the effect of test preparation on student performance on standardized examinations but the results from these studies have not been conclusive.

DESCRIPTIVE STATISTICS

We use cross-sectional data from the Saber11 examination administered in the second semester of 2012. Even though the results of the examinations in all subjects are available, we restricted the analysis to the global score and the scores in mathematics and language. However, our analysis could be extended to study other subjects that are also tested in the Colombian High School exit examination. We analyze the global score because Universities and scholarship-granting institutions in Colombia use it for their admissions and scholarship selection processes. The results in mathematics and language are also reported because these two subjects capture two essential competencies: quantitative ability and verbal ability. Additionally, these two subjects have been widely considered in previous studies about the effects of preparation.

The availability of a rich database provided by ICFES with information about the students' socioeconomic background, schools' characteristics, preparatory activities and Saber11 results opens new possibilities for the analysis of the relationship between test preparation activities and levels of performance. We considered several family and individual control variables. At the individual level, we included two variables in the analysis: gender and age. The variables we considered to account for family characteristics are parental education (mother's/father's highest schooling level), household economic status[6] (strata), whether the mother is a stay at home mother, whether the household has an internet connection and area of residence (urban or rural). Two additional variables were included to account for school characteristics: the type of school (public or private) and the school's shifts (full-time, morning only, afternoon only, evening only, or Saturday only).

As for the variables to study the effects of preparation, we considered whether the student participated in a preparatory activity or not, the institution where preparatory activities took place (at school with one of the school's teachers, at school with external faculty, at a University or a private educational institution commonly known as pre-ICFES institute), and the time dedicated to these activities (measured in months).

The original database has information from 549.823 students. However, to analyze a more homogeneous group, we restricted our analysis to students in the age range between 14 and 21 years old (92.6% of the complete sample). We also did not include in the analysis observations that have missing values for the key variables used in the regressions. Our final dataset contained information of 424.589

---

[6] The Colombian government divides the population into six socioeconomic strata, with 1 being the poorest and 6 being the richest.

students (77% of the complete sample). Table 1a shows the descriptive statistics of this final (clean) database[7].

Table 1a.
Descriptive Statistics

Variable	Total Sample		No Preparation		Preparation		Difference	SE	(p-value)
	Mean	SD	Mean	SD	Mean	SD			
Global score	254.06	43.4	248.79	39.29	255.98	44.65	7.189	0.150	0.000
Language	50.91	10.04	49.82	9.397	51.306	10.24	1.486	0.035	0.000
Mathematics	50.885	10.05	49.952	9.355	51.225	10.27	1.273	0.035	0.000
Female	0.549	0.498	0.511	0.500	0.563	0.496	0.052	0.002	0.000
Age	16.832	1.078	16.961	1.12	16.785	1.059	-0.176	0.004	0.000
Maximum education parents									
None	0.009	0.097	0.011	0.102	0.009	0.095	-0.002	0.000	0.000
Incomplete elementary	0.117	0.322	0.123	0.329	0.115	0.319	-0.008	0.001	0.000
Elementary	0.149	0.356	0.157	0.364	0.146	0.353	-0.011	0.001	0.000
High school dropout	0.159	0.366	0.169	0.375	0.156	0.363	-0.013	0.001	0.000
High school	0.296	0.456	0.310	0.462	0.291	0.454	-0.019	0.002	0.000
Incomplete associate's degree	0.019	0.135	0.019	0.136	0.018	0.135	-0.000	0.000	0.499
Associate's degree	0.064	0.244	0.058	0.234	0.066	0.248	0.008	0.001	0.000
Some college	0.026	0.159	0.023	0.150	0.027	0.162	0.004	0.001	0.000
College	0.111	0.314	0.086	0.281	0.120	0.325	0.034	0.001	0.000
Graduate degree	0.030	0.172	0.021	0.143	0.034	0.181	0.013	0.001	0.000
Doesn't know	0.020	0.138	0.023	0.151	0.018	0.133	-0.005	0.000	0.000

(*Continued*)

---

[7] We also run the descriptive statistics for the students excluded from the sample due to missing values in the variable capturing preparation. Socio-economic characteristics are less favorable in comparison to the entire sample (and also compared to the control and treatment group). Most of the excluded students belong to strata 1. have parents with lower levels of education and study in partial shifts. They also have lower performance in all subjects. These students are more likely to belong to the group that did not receive preparation (or cannot afford one).

Table 1a.
Descriptive Statistics

Variable	Total Sample		No Preparation		Preparation		Difference	SE	(p-value)
	Mean	SD	Mean	SD	Mean	SD			
Mother at home	0.523	0.499	0.528	0.499	0.521	0.500	-0.008	0.002	0.000
Household internet	0.464	0.499	0.429	0.495	0.476	0.499	0.047	0.002	0.000
Strata 1	0.384	0.486	0.383	0.486	0.385	0.486	0.001	0.002	0.393
Strata 2	0.368	0.482	0.386	0.487	0.361	0.480	-0.024	0.002	0.000
Strata 3	0.194	0.395	0.192	0.394	0.194	0.396	0.002	0.001	0.083
Strata 4	0.038	0.191	0.028	0.166	0.041	0.199	0.013	0.001	0.000
Strata 5	0.012	0.107	0.008	0.089	0.013	0.112	0.005	0.000	0.000
Strata 6	0.005	0.070	0.003	0.054	0.006	0.074	0.003	0.000	0.000
Urban	0.791	0.406	0.792	0.406	0.791	0.407	-0.001	0.001	0.399
Public school	0.757	0.429	0.775	0.418	0.751	0.433	-0.024	0.001	0.000
School hours									
Full-time	0.287	0.452	0.240	0.427	0.304	0.460	0.065	0.002	0.000
Part-time morning	0.465	0.499	0.458	0.498	0.468	0.499	0.010	0.002	0.000
Part-time evening	0.053	0.224	0.093	0.291	0.038	0.192	-0.055	0.001	0.000
Part-time Saturday	0.028	0.164	0.048	0.214	0.020	0.140	-0.028	0.001	0.000
Part-time afternoon	0.167	0.373	0.161	0.368	0.169	0.375	0.008	0.001	0.000
Observations	424.589		113.33		311.259				

Source: Authors' calculations.

Column 1 of *Table 1a* shows that the global score average is 254 points, and the standard deviation is 43. The average score in the math and language examinations is 51 points, with a standard deviation of 10. In the clean dataset, the average global score is four points higher in comparison to the original dataset. In the case of the clean dataset math and language exams, the average score is one point over the original distribution. This implies that those students who were dropped from the sample had lower scores in the examination. *Column 1* also presents the descriptive statistics of the variables that capture the family background, personal and school characteristics. In general, the students in the sample can be

characterized by the following traits: they live in urban areas (79%), belong to a low-income family (75% correspond to strata 1 and 2), attend a public school (76%), and most of their parents did not receive a college education (73%). In addition, there is a higher percentage of girls than boys (55% girls) in the sample, and the average age is 16.8 years old.

Columns 2 and 3 of *Table 1a* present a comparison of the characteristics of those students who participated in a preparatory activity for the Saber11 exam (treatment group) and those who did not (comparison group). There are systematic differences. If we compare the test scores by treatment (preparation), the average outcome in the math, language and global categories for those students who were treated slightly exceeds the average of those students who did not receive the treatment. The differences are 7.2. 1.5 and 1.3 points, respectively (0.16. 0.15 and 0.13 in a normalized Z-score), and they are all significantly different from zero (*column 4*). However, it is unclear whether the differences between the two groups are caused by the preparatory activities, the time spent on those activities or because the two groups differ in some relevant observed characteristics that affect their outcomes on the exam.

In fact, by examining the observed characteristics, it is possible to determine that those students who were treated tend to have more-favorable socioeconomic characteristics in comparison to their counterparts. The descriptive statistics (*Columns 2 and 3 - Table 1a*) show that a higher proportion of the treated students have the following characteristics: they spend more time at school (full-time, all morning or all afternoon), belong to a higher socio-economic stratum (strata 4. 5 and 6), and their parents have a higher level of education (high school, college and graduate degree). All these differences are statistically significant (*Column 4*). Other variables, such as the share of students residing in urban areas (75%), are similar between the two groups.

As part of the analysis, we also considered where the preparatory activities took place and the different durations. *Table 1b* presents descriptive information about the places where the preparation was conducted, by duration. Students are unevenly divided among the different places and the different durations. As reported in *Table 1b*, 71% of the students took a course at school (45% with school staff and 26% with external staff), 9% took a course at a Pre-ICFES institute, 1% took a preparatory course at a local University, and 18% of the students enrolled in preparatory activities at more than one place. There are also significant differences in duration. For instance, according to our data, 61% of the students who enrolled in a Pre-ICFES course participated in that activity for more than three months. On the other hand, 71% of the students who took a course at school, taught by school faculty, reported duration of fewer than three months.

Table 1b.
Preparation by institution and duration

Preparatory Activity - Institution	Less than 1 month	Between 1 & 3 months	Between 3 & 6 months	More than 6 months	Total	% students by activity
School - School staff	51,263	47,729	24,187	16,262	139,441	45%
(% by duration)	37%	34%	17%	12%	100%	
School - External faculty	16,866	27,565	26,170	10,471	81,072	26%
(% by duration)	21%	34%	32%	13%	100%	
Course (Pre-ICFES Institution)	3,848	7,545	12,357	5,554	29,304	9%
(% by duration)	13%	26%	42%	19%	100%	
Course (University)	885	1,108	1,141	392	3,526	1%
(% by duration)	25%	31%	32%	11%	100%	
Several	8,939	17,351	15,341	14,340	55,971	18%
(% by duration)	16%	31%	27%	26%	100%	
Total	81,802	101,299	79,197	47,020	309,318	100%

Source: Authors' calculations.

Differences between the treatment and the comparison groups cast doubts on the validity of the estimates obtained using a linear method that does not compare comparable students. Therefore, in addition to OLS, we also use a Propensity Score Matching (PSM) methodology.

METHODOLOGIES

In this research, two econometric methods are used for the data analysis: 1) cluster-robust ordinary least squares, and 2) propensity score matching. The application of both methodologies will be explained in more detail below.

Ordinary Least Squares

To obtain a first estimate of the relationship between participating in test preparation activities and the score obtained in the Saber11 exit examination, we use an ordinary least squares regression model. However, to take into consideration the hierarchical structure of the dataset (i.e., students are enrolled in different schools), we employ a cluster-robust version of the model that corrects the

standard errors for within-cluster error correlation. Specifically, we considered student-level cross-sectional data with school clustering[8].

Propensity Score Matching

The matching process must be performed considering a full range of variables across which the students in the treatment group and the students in the control group might differ to eliminate potential bias. However, considering a full range of variables can be problematic in terms of dimensionality (Heinrich, Maffioli, & Vazquez., 2010). Therefore, it is necessary to make use of the propensity score matching (PSM) technique. Using PSM, we can tackle the dimensionality problem by defining a propensity score that allows us to carry out the analysis using a single dimension. In other words, instead of trying to match all values of the variables, by using PSM, we compare students based on propensity scores (Heinrich et al., 2010).

Two conditions should be met to implement the propensity score matching methodology. First, the variables on which the treatment and control groups differ have to be observable (the *conditional independence* or *unconfoundedness* assumption). Our database, which contains a complete set of information about the students who take the Saber11 examination, can help us meet that condition. Second, the *common support* or *overlap condition* should be met (Heinrich et al., 2010), which is visually tested.

The first step to implement the propensity score matching method is to estimate the propensity score. In this case, the propensity score captures the probability of participating in Saber11 preparatory activities. Considering that the treatment status is dichotomous, a probit model is estimated. In this case, we estimate the following model:

$$PREPARATION = f(GENDER, PARENTAL\ EDUCATION, PUBLIC\ SCHOOL,$$
$$SCHOOL\ LOCATION, MEETING\ TIME, SOCIOECONOMIC\ STRATA) \tag{2}$$

The variable choice is guided by the conditional independence assumption that requires that the outcome variable be independent of treatment conditional on the propensity score. In other words, only variables that influence the participation decision and the outcome variable simultaneously, and those variables that are unaffected by participation were included in the analysis (Caliendo & Kopeinig, 2008). The choice of the matching variables was also based on the theoretical framework discussed in Section V and the previous empirical studies cited in the literature review. *Table 2* provides a description of the variables included in the model used to estimate the probability of participating in Saber11 preparatory activities. *Table 3* reports Probit estimates for the propensity score. Most of

---

[8] Failure to control for within-cluster error correlation can lead to misleadingly small standard errors and, therefore, misleadingly narrow confidence intervals, large t-statistics and low p-values.

the variables considered in the analysis are significantly related to the probability of participating in preparatory activities. The fact that those variables are significant, and the signs of the coefficients are as expected is a good sign of result suitability.

Table 2.
Description of the Variables Included in the Probit Model

	Description
Gender	Specified as a dummy variable for female students (Female = 1).
Age	This variable is included as a continuous variable.
Parental Education	We include dummy variables to measure the highest level of education achieved by the parents of the student: incomplete elementary, elementary, incomplete High School, High School, incomplete associate's degree, associate's degree, incomplete college, college, and graduate degree. The omitted category is no education.
Public School	Specified as a dummy variable (Public School = 1). Approximately 76% of the students attend public schools.
School's Meeting Time	We include dummy variables to take into account that some schools in Colombia have different meeting times: morning, afternoon, night, Saturday. The omitted category is full time.
Stay at Home Mother	About half of the mothers of the students in our sample are stay at home mothers. We include mothers at home to control for the fact that these mothers may be spending more time monitoring their children´s duties, including pre-test preparation. Specified as a dummy variable (The students' mother stays at home = 1).
Household has Internet Service	Specified as a dummy variable (Internet service = 1).
Socioeconomic Strata	Proxy for family income. Law 142 of 1994 mandates the use of stratification to classify Colombian population in distinct groups (strata) with similar social and economic characteristics. Housing characteristics are the main criteria used for this classification. There could be six strata: level one is lower-low, two is low, three is upper-low, four is medium, five is medium-high, and six is high. The omitted category in our empirical exercise is Strata 6.

Source: Authors' calculations.

Table 3.

Probit on Preparation for the Saber11 Examination – Complete

VARIABLES	(1) dy/dx
Female=1	0.0431***
	(0.00151)
Maximum education parents: Incomplete elementary	0.0193**
	(0.00765)
Maximum education parents: Elementary	0.0181**
	(0.00761)
Maximum education parents: High school dropout	0.00490
	(0.00772)
Maximum education parents: High school	0.000594
	(0.00769)
Maximum education parents: Incomplete associate's degree	0.00107
	(0.00933)
Maximum education parents: Associate's degree	0.0141**
	(0.00805)
Maximum education parents: Some college	0.0212***
	(0.00872)
Maximum education parents: College	0.0334***
	(0.00775)
Maximum education parents: Graduate degree	0.0450***
	(0.00855)
Maximum education parents: Doesn't know	-0.019*
	(0.00936)
Strata 1	-0.00864
	(0.0130)
Strata2	-0.0150
	(0.0130)
Strata3	-0.0281**
	(0.0131)
Strata4	-0.0140
	(0.0132)

(*Continued*)

Table 3.
Probit on Preparation for the Saber11 Examination – Complete

VARIABLES	(1) dy/dx
Strata5	-0.0164
	(0.0145)
Age	-0.0136***
	(0.00215)
Public school	-0.0133***
	(0.00209)
Meeting time: morning	-0.042***
	(0.00188)
Meeting time: night	-0.248***
	(0.00384)
Meeting time: Saturday	-0.235***
	(0.00508)
Meeting time: afternoon	-0.039***
	(0.00251)
Stay at home mother	0.00451***
	(0.00160)
Household has internet service	0.017***
	(0.00183)
Observations	**424.589**
Pseudo R-squared	**0.0393**

Standard errors in parentheses
* p<0.10, ** p<0.05. *** p<0.01
Source: Authors' calculations.

A fundamental step to investigate the validity of the PSM estimation is to verify the common support condition (Heinrich et al., 2010). This condition warrants that units with the same characteristics have a positive probability of being either participants or nonparticipants. For this empirical exercise, we check the condition of common support between treatment and comparison groups by visually inspecting the propensity score distributions. *Figure 1* presents the kernel distribution and the common support area for those students who got prepared for the exam [treated] and those who did not [untreated]. As the graph shows, there is a high degree of overlap in the propensity scores of the treated and untreated units. In other words, the characteristics observed in the treatment group are also observed in the control

group (Caliendo and Kopeinig, 2008). The common support condition guarantees that each treated student has a matching unit in the group of untreated students.

Figure 1.
Kernel distribution and common support area across the 2 groups – Complete Sample

Kernel density estimate

kernel = epanechnikov, bandwidth = 0.0070
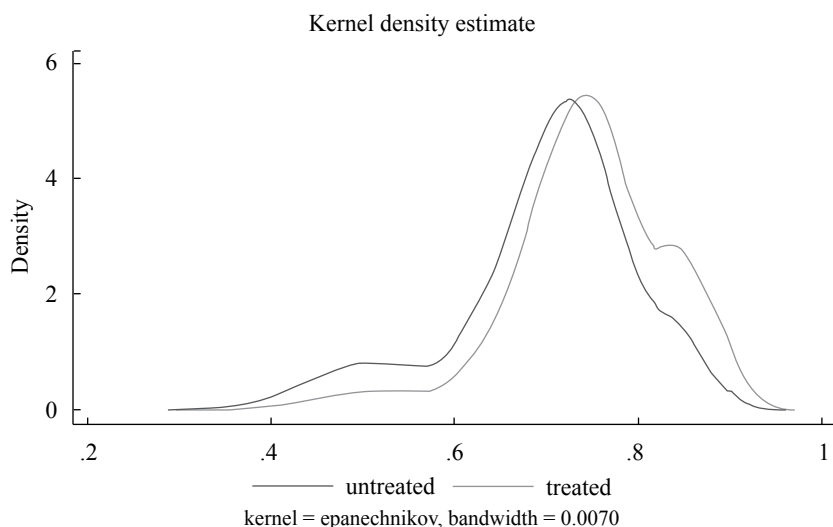
Source: Authors' calculations.

Table 4 presents evidence of covariate balance after matching. T-statistics for the equality of covariate means demonstrate that those students who participated in a preparatory activity and those who did not have very similar observable characteristics within each block in all cases. There were no significant differences in the means of the covariates between the treated and the control groups. The bias before and after matching, and the change in this bias, is also reported for each variable. We conclude that the balancing property is satisfied.

Table 4.
Balancing tests - Nearest Neighbors

Variable	Unmat-ched	Mean		%bias	%reduct	t-test	
	Matched	Treated	Control		|bias|	t	p>|t|
Female=1	U	0.56257	0.51057	10.4	98.3	30.15	0
	M	0.56255	0.56359	-0.2		-0.69	0.492

(*Continued*)

Table 4.
Balancing tests - Nearest Neighbors

Variable	Unmatched / Matched	Mean		%bias	%reduct |bias|	t-test	
		Treated	Control			t	p>|t|
Maximum education parents: Incomplete elementary	U	0.1151	0.12316	-2.5	97.6	-7.23	0
	M	0.1151	0.11532	-0.1		-0.24	0.812
Maximum education parents: Elementary	U	0.14621	0.1569	-3	90.2	-8.66	0
	M	0.14621	0.14716	-0.3		-1.17	0.241
Maximum education parents: High school dropout	U	0.15585	0.16903	-3.6	88.5	-10.38	0
	M	0.15586	0.15738	-0.4		-0.21	0.831
Maximum education parents: High school	U	0.2907	0.30987	-4.2	87.6	-12.11	0
	M	0.29071	0.29308	-0.5		-0.29	0.772
Maximum education parents: Associate's degree	U	0.06581	0.0579	3.3	91.9	9.33	0
	M	0.06581	0.06517	0.3		1.02	0.306
Maximum education parents: Some college	U	0.02687	0.02314	2.4	81.8	6.78	0
	M	0.02687	0.02619	0.4		0.13	0.895
Maximum education parents: College	U	0.11988	0.08634	11	99.5	30.82	0
	M	0.11988	0.12005	-0.1		-0.2	0.842
Maximum education parents: Graduate degree	U	0.03398	0.02075	8.1	99.1	22.2	0
	M	0.03395	0.03408	-0.1		-0.24	0.807
Maximum education parents: Doesn't know	U	0.01807	0.0235	-3.8	74.8	-11.3	0
	M	0.01808	0.01671	1		0.13	0.898
Strata 1	U	0.38453	0.38308	0.3	-24.7	0.85	0.393
	M	0.38454	0.38633	-0.4		-1.46	0.145
Strata 2	U	0.36148	0.38579	-5	97.5	-14.54	0
	M	0.36149	0.36209	-0.1		-0.5	0.62
Strata 3	U	0.19438	0.19201	0.6	89	1.73	0.083
	M	0.19439	0.19412	0.1		0.26	0.795
Strata 4	U	0.04125	0.02825	7.1	91.7	19.65	0
	M	0.04123	0.04018	0.6		0.14	0.891
Strata 5	U	0.01282	0.00791	4.8	85.9	13.27	0
	M	0.01281	0.01212	0.7		0.29	0.773

(*Continued*)

Table 4.

Balancing tests - Nearest Neighbors

Variable	Unmat-ched	Mean		%bias	%reduct	t-test	
	Matched	Treated	Control		|bias|	t	p>|t|
Age	U	16.529	16.71	-16.5	99.5	-48.2	0
	M	16.529	16.528	0.1		0.34	0.737
Stay at home mother	U	0.52059	0.52818	-1.5	88.2	-4.38	0
	M	0.52058	0.52148	-0.2		-0.71	0.477
Household has internet service	U	0.47614	0.42883	9.5	98	27.37	0
	M	0.47614	0.47519	0.2		0.75	0.456
Public school	U	0.75057	0.77504	-5.8	89.5	-16.45	0
	M	0.75058	0.75315	-0.6		-0.99	0.321
Meeting time: morning	U	0.46791	0.45769	2.1	95.7	5.91	0
	M	0.46793	0.46747	0.1		0.35	0.726
Meeting time: night	U	0.03834	0.09316	-22.3	99.6	-70.96	0
	M	0.03834	0.03856	-0.1		-0.44	0.659
Meeting time: Saturday	U	0.02002	0.04805	-15.5	98.9	-49.55	0
	M	0.02002	0.0197	0.2		0.88	0.378
Meeting time: afternoon	U	0.16927	0.16121	2.2	82.9	6.23	0
	M	0.16927	0.1679	0.4		1.45	0.147

Source: Authors' calculations.

RESULTS

This section presents the results of the regression analysis on the relationship between participation in preparatory activities for the Saber11 examination and the scores obtained in the Saber11 examination.

The results obtained using ordinary least squares are presented in *Table 5*. The specification without control variables, or naïve specification (*Columns 1. 4 and 7*), show that the effect of the preparation is positive and significant. The results reveal that the effect of the preparation is 7.2. 1.5 and 1.3 points on the results of the global examination, the language and the mathematics examinations, respectively. *Columns 2. 3. 5. 6 and 8* in *Table 5* present the results of the specifications that control for observable characteristics of the individuals and the schools. The inclusion of these variables attenuates the effect of preparatory activities. This implies that preparatory activities are highly correlated with observable individual

and school characteristics. The results for these specifications indicate that participation in preparatory activities increases the global score by 3.13 points, the language score by 0.58 points and the math score by 0.59 points. The magnitude of this effect is modest if we consider that the standard deviation of the global, the language and the mathematics scores is 43.4. 10.04 and 10.05 points, respectively. In other words, the preparatory activities increase examination scores by approximately 0.058 to 0.07 standard deviations (SD). It should be noted that our measure of preparation includes activities that are offered both at school and out of school. Approximately 45% of the students were prepared at school by in-house staff, while only 9% at a pre-ICFES institute. The pre-ICFES institutes have personnel specialized in these tests, while most schools don't, which could explain at least in part why our results seem to be modest.

Our results can be contrasted with other results reported in the literature to get a better sense of the effect. Using an instrumental variables approach, Ryu and Kang (2013) find that a 10 percent increase in private tutoring expenditures raises the average overall score by 0.03 standard deviations (S.D). It should be noted that these researchers find an even smaller effect using a matching strategy (Ryu & Kang, 2013). Zhang and Liu (2016), estimate an OLS model and find a positive effect of large class mathematics tutoring of 0.248 S.D. In the case of small group language (Chinese) tutoring, they find a positive effect of 0.293 S.D. In sum, our results show a more substantial effect in comparison to Ryu and Kang's (2013) findings but are very modest in comparison to Zhang and Liu's (2016). Nevertheless, it is worth highlighting that the test preparation activities evaluated in these studies differ to those activities evaluated in our study.

Another way to have a better sense of the magnitude of the effect of the preparatory activities on exam performance is to compare the coefficient of the variable that captures preparation with the coefficients of other variables that were found to have a sizeable effect on academic performance in previous studies for Colombia. *Table 5* reports the coefficients for gender, socioeconomic status and the nature of the institution (private or public). The results reveal that being female, belonging to socioeconomic strata 1 and 2. and studying in a public school have a negative and significant effect on the scores in the three examinations. These findings are in line with the results found in other studies for Colombia (Chica, Galvis, & Ramírez, 2011; Núñez, Steiner, Cadena & Pardo; 2002; Tobón, Posada, & Ríos, 2009; Gómez Soler 2016; 2017). In the most comprehensive specification (*Columns 3. 6 and 9*), being a woman reduces the global score by 8 points (0.18 SD). Hence, preparation activities seem to have a minimal effect if we compare the results discussed above with the results obtained for the preparation (treatment) variable.

Table 5.
OLS estimates of the effect of preparation on Saber 11° performance*

Variables	Global score			Language			Math		
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Preparation	7.1892***	4.6396***	3.1310***	1.4864***	0.8999***	0.5810***	1.2732***	0.8751***	0.5932***
	[0.3737]	[0.2483]	[0.2344]	[0.0762]	[0.0515]	[0.0496]	[0.0759]	[0.0546]	[0.0519]
Female		-8.1767***	-8.0506***		-0.0635	-0.0413		-3.3726***	-3.3501***
		[0.2323]	[0.2249]		[0.0470]	[0.0456]		[0.0517]	[0.0505]
Strata 1 & 2 (Yes==1)		-14.5721***	-13.2027***		-2.8016***	-2.5542***		-2.6305***	-2.3939***
		[0.4027]	[0.3596]		[0.0779]	[0.0716]		[0.0868]	[0.0781]
Public school			-5.5502***			-0.9796***			-0.9162***
			[0.5839]			[0.1159]			[0.1230]
Constant	248.7932***	366.5173***	358.4593***	49.8200***	72.8758***	70.8693***	49.9518***	75.0878***	73.5758***
	[0.3814]	[1.7447]	[1.8003]	[0.0805]	[0.3821]	[0.3866]	[0.0742]	[0.3960]	[0.4116]
Individual characteristics	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
School characteristics	No	No	Yes	No	No	Yes	No	No	Yes
R-Square	0.005	0.275	0.29	0.004	0.2	0.213	0.003	0.203	0.213
Number of observations	424589	424589	424589	424589	424589	424589	424589	424589	424589

Notes:
*SABER Global Score is calculated as a weighted average of the five subjects. Individual characteristics are age, highest level of education of parents (references level is none education) and area of residence (reference level is rural area). School characteristics is meeting time (references level is full-time). Standard errors in square brackets.
* $p<0.10$, ** $p<0.05$. *** $p<0.01$.
Source: Authors' calculations.

Table 6.

Impact of Preparation – Propensity Score Matching

	Nearest Neighbor			**5 Nearest Neighbors**			**Caliper Matching 0.001**		
SUBJECT	**Diff.**	**S.E.**	**z**	**Diff.**	**S.E.**	**z**	**Diff.**	**S.E.**	**z**
Global Score	4.316***	0.350	12.32	3.939***	0.171	23.04	4.326***	0.296	14.60
Mathematics	0.773***	0.079	9.83	0.763***	0.042	18.06	0.775***	0.084	9.23
Language	0.865***	0.054	16.11	0.749***	0.047	15.91	0.867***	0.066	13.04
Observations On Support	424,559			424,559			424,378		
Observations Off Support	30			30			211		
Observations	424,589			424,589			424,589		

Note: each column reports the matching estimator with a different matching algorithm (1) nearest neighbor matching using 1 nearest neighbor (2) nearest neighbor matching using 5 nearest neighbors (3) caliper matching with a caliper of 0.001.
The standard errors reported are bootstrapped standard errors.
* p<0.10, ** p<0.05. *** p<0.01.
Source: Authors' calculations.

The average treatment effect on those students who participated in preparatory activities (ATT), which was estimated using propensity score matching (PSM), is presented in *Table 6*. In line with the results that were obtained using cluster robust ordinary least squares, the estimates obtained with PSM show a positive and significant effect of preparation on the global score. Specifically, the results obtained using the five nearest neighbors algorithm indicate that participation in preparatory activities is associated with an increase in the score of 3.94 points in the global score, 0.75 points in language, and 0.76 in the math test. As shown in *Table 6.* the results are robust to the matching algorithms employed for the analysis. It should be highlighted that even though the effect obtained using the matching technique is larger than the effect using cluster robust OLS, it is still fairly small.

Heterogeneous Effects

An interesting question is whether preparation activities have a greater impact among the poorest, female, and public-school students, as these variables are found to have a large and negative effect on tests scores. It might be the case that preparatory activities can help close the gap in the scores of these subgroups. *Tables 7a, 7b* and *7c* present the results of our heterogeneous effects regression by gender, socioeconomic status and public schools. The interactions of the preparation variable with the dummy variables for public schools and low-income students are

negative and statistically significant. These results suggest that preparatory activities are less useful for low-income students attending public schools by approximately 5 points in the global score compared to their peers (0.12 SD). In contrast, the results for the interaction between the gender dummy variable (female=1) and the preparation variable show that preparation is more beneficial for female students in comparison to male students, as it increases their global scores by approximately 0.75 points (0.017 SD).

The results also show that, on average, a male student who attends a private institution and belongs to the highest socioeconomic status can increase his global score by 11 points by participating in preparatory activities. This corresponds to an increase of 0.25 SD.

Preparatory Activities and the Effects of Duration

The data we employed in this study contains information about both participation in preparatory activities and the institution where these activities took place. This is especially important because the quality and the duration of the activities offered by each of these institutions might be very different. *Table 8* presents the results by institution. After controlling for individual and school characteristics (*Table 8 column 3*), we find that enrolling at a Pre-ICFES institution increases the global score by 13.42 points (0.3 SD). Participating in a preparatory course offered by a university increases the global score by 7.66 points (0.18 SD). Meanwhile, preparatory activities offered at school increase the global score by 5.89 (0.14 SD) points if taught by external faculty and by only 1 point (0.02 SD) if taught by school staff. The effect on language and math is very similar, if not identical.

We also controlled for the time spent on preparatory activities. The duration was measured in the following categories: less than one month, between 1 and 3 months, between 3 and 6 months, and more than 6 months. However, as students could enroll in preparatory courses in more than one place, they could also report a different duration for each place. To define a unique duration for each student, we use the maximum duration of all the preparatory activities reported by a student. *Table 9* reports the results of the effect of course duration on test scores. The results show that preparatory activities that last longer than three months have a significant and broad impact on test scores. More specifically, we found an effect of between 0.18 and 0.28 SD depending on the subject. On the other hand, activities that last less than one month have a negative effect. However, the magnitude of this negative effect is quite small, approximately 0.07 SD.

Table 7a.
OLS estimations of the heterogeneity of preparation on the Global Score

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Preparation	7.5764***	19.8327***	15.2200***	20.7231***	2.8766***	8.4904***	8.6054***	11.0459***
	[0.4571]	[1.0913]	[0.8569]	[1.1859]	[0.2871]	[0.6749]	[0.5963]	[0.8403]
Preparation & Female	0.0436			0.7336*	0.485			0.7465**
	[0.4805]			[0.3963]	[0.3296]			[0.3307]
Preparation & Public school		-17.1863***		-12.4690***		-6.9123***		-5.1826***
		[1.1337]		[0.9269]		[0.7161]		[0.6978]
Preparation & Group			-11.5211***	-6.4682***			-7.1721***	-5.6181***
			[0.8518]	[0.7172]			[0.5885]	[0.5493]
Female	-7.9188***			-7.3748***	-8.4048***	-8.0427***	-8.0486***	-8.5882***
	[0.4687]			[0.3971]	[0.3050]	[0.2242]	[0.2245]	[0.3135]
Public school		-10.4658***		-3.4180***	-5.5508***	-0.4477	-5.4428***	-1.6412**
		[1.2333]		[1.0161]	[0.5839]	[0.7244]	[0.5809]	[0.7299]
Low-income (strata 1.2)			-24.6553***	-23.4921***	-13.2026***	-13.1428***	-7.9496***	-9.0428***
			[0.8278]	[0.7087]	[0.3596]	[0.3564]	[0.4983]	[0.4877]
Constant	252.8363***	256.9046***	267.7501***	273.2702***	358.6414***	354.1908***	354.1991***	352.2020***
	[0.4329]	[1.1774]	[0.8444]	[1.2435]	[1.7920]	[1.8093]	[1.8613]	[1.8644]
Individual characteristics	No	No	No	No	Yes	Yes	Yes	Yes
School characteristics	No	No	No	No	Yes	Yes	Yes	Yes
R-Square	0.014	0.064	0.117	0.14	0.29	0.291	0.291	0.292
Number of observations	424,589	424,589	424,589	424,589	424,589	424,589	424,589	424,589

Individual characteristics are age, highest level of education of parents (reference level is no education) and area of residence (reference level is rural area). School characteristic is meeting time (reference level is full-time). Standard errors in [].
* $p<0.10$, ** $p<0.05$. *** $p<0.01$.
Source: Authors' calculations.

Table 7b.
OLS estimations of the differential impact of preparation on the Language Score by Subgroup

Variables	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Preparation	1.4599***	3.8957***	2.9068***	3.8734***	0.5090***	1.5198***	1.5347***	1.9382***
	[0.0907]	[0.2113]	[0.1641]	[0.2230]	[0.0605]	[0.1348]	[0.1166]	[0.1625]
Preparation & Female	0.0502			0.1808**	0.1373*			0.1830**
	[0.1015]			[0.0864]	[0.0737]			[0.0741]
Preparation & Public school		-3.2757***		-2.4004***		-1.2109***		-0.9122***
		[0.2219]		[0.1845]		[0.1445]		[0.1450]
Preparation & Group			-2.0583***	-1.0957***			-1.2495***	-0.9774***
			[0.1647]	[0.1450]			[0.1168]	[0.1137]
Female	-0.0334			0.0768	-0.1415**	-0.0399	-0.0409	-0.1736**
	[0.1008]			[0.0866]	[0.0682]	[0.0455]	[0.0455]	[0.0696]
Public school		-2.0183***		-0.6232***	-0.9798***	-0.0858	-0.9609***	-0.2918*
		[0.2469]		[0.2031]	[0.1159]	[0.1480]	[0.1155]	[0.1503]
Group (Strata 1&2 =1)			-5.0347***	-4.8609***	-2.5541***	-2.5437***	-1.6390***	-1.8304***
			[0.1645]	[0.1451]	[0.0716]	[0.0711]	[0.1030]	[0.1030]
Constant	49.8371***	51.3843***	53.6911***	54.0012***	70.9209***	70.1216***	70.1272***	69.7942***
	[0.0878]	[0.2332]	[0.1669]	[0.2377]	[0.3853]	[0.3876]	[0.3978]	[0.3970]
Individual characteristics	No	No	No	No	Yes	Yes	Yes	Yes
School characteristics	No	No	No	No	Yes	Yes	Yes	Yes
R-Square	0.004	0.044	0.086	0.096	0.213	0.214	0.214	0.214
Number of observations	424,589	424,589	424,589	424,589	424,589	424,589	424,589	424,589

Individual characteristics are age, highest level of education of parents (references level is none education) and area of residence (reference level is rural area). School characteristics is meeting time (references level is full-time). Standard errors in square brackets
* $p<0.10$, ** $p<0.05$. *** $p<0.01$.
Source: Authors' calculations.

Table 7c.
OLS estimations of the differential impact of preparation on the Mathematics Score by Subgroup

Variables	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Preparation	1.4565***	3.7836***	2.8727***	4.1036***	0.5619***	1.7071***	1.6913***	2.2286***
	[0.0968]	[0.2220]	[0.1793]	[0.2510]	[0.0662]	[0.1461]	[0.1310]	[0.1853]
Preparation & Female	-0.0233			0.1109	0.0596			0.1133
	[0.1025]			[0.0883]	[0.0775]			[0.0773]
Preparation & Public school		-3.4012***		-2.4924***		-1.4367***		-1.0937***
		[0.2308]		[0.1928]		[0.1557]		[0.1506]
Preparation & Group			-2.2805***	-1.2916***			-1.4386***	-1.1095***
			[0.1788]	[0.1508]			[0.1305]	[0.1215]
Female	-3.2724***			-3.1803***	-3.3936***	-3.3484***	-3.3497***	-3.4312***
	[0.0965]			[0.0850]	[0.0703]	[0.0503]	[0.0504]	[0.0714]
Public school		-1.7354***		-0.3867*	-0.9162***	0.1444	-0.8946***	-0.0923
		[0.2402]		[0.2032]	[0.1230]	[0.1517]	[0.1223]	[0.1519]
Group (Strata 1&2 =1)			-4.4479***	-4.2531***	-2.3939***	-2.3815***	-1.3402***	-1.5718***
			[0.1658]	[0.1431]	[0.0781]	[0.0774]	[0.1075]	[0.1048]
Constant	51.6226***	51.2968***	53.3717***	55.1454***	73.5981***	72.6886***	72.7212***	72.2839***
	[0.0887]	[0.2294]	[0.1688]	[0.2538]	[0.4103]	[0.4149]	[0.4233]	[0.4258]
Individual characteristics	No	No	No	No	Yes	Yes	Yes	Yes
School characteristics	No	No	No	No	Yes	Yes	Yes	Yes
R-Square	0.03	0.04	0.075	0.109	0.213	0.214	0.214	0.214
Number of observations	424,589	424,589	424,589	424,589	424,589	424,589	424,589	424,589

Individual characteristics are age, highest level of education of parents (references level is none education) and area of residence (reference level is rural area). School characteristics is meeting time (references level is full-time). Standard errors in square brackets.
* p<0.10, ** p<0.05. *** p<0.01.
Source: Authors' calculations.

Table 8.
OLS estimates of the effect of preparation type on Saber 11° performance*

Variables	Global score*			Language				Math	
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
School - School staff	-0.4204	2.1082***	1.0018***	-0.0330	0.4092***	0.1751***	-0.1522	0.3972***	0.1896***
	[0.4748]	[0.3149]	[0.2933]	[0.0953]	[0.0639]	[0.0605]	[0.0969]	[0.0692]	[0.0652]
School - External faculty	13.3681***	7.7200***	5.8951***	2.8629***	1.6292***	1.2400***	2.3676***	1.3914***	1.0495***
	[0.5438]	[0.3723]	[0.3647]	[0.1097]	[0.0759]	[0.0748]	[0.1100]	[0.0818]	[0.0803]
Course (Pre-Icfes)	31.8248***	15.3540***	13.4197***	6.3686***	2.9801***	2.5746***	5.8696***	2.8741***	2.5199***
	[0.7334]	[0.4733]	[0.4772]	[0.1413]	[0.0953]	[0.0963]	[0.1579]	[0.1095]	[0.1107]
Course (University)	23.2892***	9.4744***	7.6588***	4.8317***	1.9424***	1.5630***	3.9596***	1.5197***	1.1826***
	[1.2548]	[0.8468]	[0.8439]	[0.2723]	[0.1942]	[0.1925]	[0.2433]	[0.1821]	[0.1824]
More than 1 course	3.8955***	1.2415***	-0.4677	0.6422***	0.0545	-0.3049***	0.7814***	0.3600***	0.0404
	[0.5643]	[0.3662]	[0.3523]	[0.1141]	[0.0755]	[0.0734]	[0.1157]	[0.0814]	[0.0788]
Female		-8.2624***	-8.1330***		-0.0797*	-0.0568		-3.3920***	-3.3689***
		[0.2324]	[0.2252]		[0.0470]	[0.0456]		[0.0517]	[0.0505]
Group (Strata 1&2 =1)		-13.8920***	-12.6067***		-2.6692***	-2.4382***		-2.5040***	-2.2827***
		[0.4096]	[0.3622]		[0.0790]	[0.0720]		[0.0885]	[0.0789]
Public School			-5.4006***			-0.9542***			-0.8889***
			[0.5863]			[0.1163]			[0.1236]
Constant	248.7932***	363.5636***	355.9466***	49.8200***	72.2859***	70.3740***	49.9518***	74.5420***	73.1130***
	[0.3814]	[1.7573]	[1.8131]	[0.0805]	[0.3845]	[0.3893]	[0.0742]	[0.3992]	[0.4147]
Individual characteristics	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
School characteristics	No	No	Yes	No	No	Yes	No	No	Yes
R-Square	0.044	0.281	0.296	0.034	0.206	0.218	0.028	0.207	0.217
Number of observations	422,384	422,384	422,384	422,384	422,384	422,384	422,384	422,384	422,384

Notes: *Saber Global Score is calculated as a weighted average of the five subjects. Individual characteristics are age, highest level of education of parents (references level is none education) and area of residence (reference level is rural area). School characteristics is meeting time (references level is full-time). Standard errors in square brackets.
* $p<0.10$, ** $p<0.05$. *** $p<0.01$.
Source: Authors' calculations.

Table 9.
OLS estimates of the effect of duration of preparation on Saber 11° performance*

Variables	Global score*			Language			Math		
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Less than 1 month	-6.4211***	-2.3483***	-2.8745***	-1.2127***	-0.4610***	-0.5808***	-1.4122***	-0.5644***	-0.6679***
	[0.3405]	[0.2496]	[0.2357]	[0.0745]	[0.0562]	[0.0544]	[0.0698]	[0.0552]	[0.0526]
Between 1 & 3 months	1.6568***	1.6977***	0.5140**	0.4731***	0.3992***	0.1336**	0.2074***	0.2908***	0.0667
	[0.3445]	[0.2608]	[0.2499]	[0.0736]	[0.0564]	[0.0550]	[0.0701]	[0.0574]	[0.0553]
Between 3 & 6 months	19.3539***	11.5658***	9.4387***	3.9287***	2.2747***	1.8331***	3.6040***	2.2346***	1.8479***
	[0.5095]	[0.3291]	[0.3218]	[0.1026]	[0.0687]	[0.0676]	[0.1040]	[0.0727]	[0.0715]
More than 6 months	23.3536***	14.2840***	11.9803***	4.4653***	2.5702***	2.1026***	4.5305***	2.9326***	2.5194***
	[0.8487]	[0.5241]	[0.5067]	[0.1607]	[0.1015]	[0.0990]	[0.1788]	[0.1178]	[0.1145]
Gender (Female=1)		-8.2913***	-8.1930***		-0.0837*	-0.0666		-3.4012***	-3.3848***
		[0.2259]	[0.2198]		[0.0461]	[0.0448]		[0.0504]	[0.0495]
Group (Strata 1&2 =1)		-13.3211***	-12.4656***		-2.5696***	-2.4175***		-2.3738***	-2.2398***
		[0.3753]	[0.3444]		[0.0737]	[0.0693]		[0.0810]	[0.0748]
Public School (Yes=1)			-3.7073***			-0.6438***			-0.5322***
			[0.5595]			[0.1118]			[0.1179]
Constant	248.7932***	363.3005***	354.8187***	49.8200***	72.2747***	70.2073***	49.9518***	74.4252***	72.8131***
	[0.3814]	[1.7119]	[1.7586]	[0.0805]	[0.3775]	[0.3810]	[0.0742]	[0.3900]	[0.4033]
Individual characteristics	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
School characteristics	No	No	Yes	No	No	Yes	No	No	Yes
R-Square	0.059	0.290	0.302	0.042	0.210	0.221	0.042	0.215	0.223
Number of observations	421,708	421,708	421,708	421,708	421,708	421,708	421,708	421,708	421,708

Notes:
*Saber Global Score is calculated as a weighted average of the five subjects. Individual characteristics are age, highest level of education of parents (references level is none education) and area of residence (reference level is rural area). School characteristics is meeting time (references level is full-time). Standard errors in square brackets.
* p<0.1. ** p<0.05. *** p<0.01.
Source: Authors' calculations.

Robustness Checks

We run robustness checks using subsamples that include additional information on the students taking the examination. First, we estimate a specification using the subset of students who took the Pre-Saber exam to control for ability differences before taking the Saber11 examination. We lost some observations, but the sample is still large (37,263 students). *Table 10a* reports the results. All the estimates are consistent but slightly smaller compared to the estimates in *Table 5*.

Second, we consider a subset of students who were randomly questioned about their educational aspirations after high school graduation. We use this variable as a proxy of motivation because it is highly likely that those students who plan to go to college are more motivated to perform better on the Saber11 exam. *Table 10b* shows that when including the motivation proxy, all the estimates are also consistent but are slightly smaller compared to previous estimates.

We estimated another specification controlling for those students who took the Saber11 examination more than once. *Table 10c* presents the results. All the estimates are consistent but are slightly smaller compared to the estimates presented in the results section. This suggests that retaking the test does not have the same effect as taking a preparation course.

The results of the robustness checks are consistent with our previous findings, where preparatory activities tended to have a modest effect. As expected, our results show that omitting variables such as previous ability and motivation could lead to an overestimation of the effect. Therefore, it is likely that the effect of the preparation is even more modest than the results with the complete sample suggested.

DISCUSSION OF THE RESULTS

Getting prepared for the Colombian exit examination is associated with a 3-point increase (0.06 of a standard deviation) in the global score of the test. On average, this result seems modest in comparison to other interventions studied in the literature (Krueger, 2003; Banerjee, Cole, Duflo, & Linden, 2007). However, this small effect could significantly affect students who are close to the cutoff point of some scholarship programs awarded by the government. For example, three points could have been the difference for approximately 2.200 low-income students who in the year 2014 were a few points away from meeting the eligibility criteria for one of the most important government programs of the last years, *Ser Pilo Paga* (SPP– which stands for Being Smart Pays Off). SPP was a merit-need-based scholarship that covered the full cost of tuition at an accredited (high-quality) university in Colombia. The Colombian government awarded this scholarship to 10.000 students every year from 2014 to 2018. for a total of 50.000 students. To be eligible, prospective students had to obtain a score above a certain threshold (310 in 2014). The program had no strings attached, but the beneficiaries had to be in good standing at their programs.

Table 10a.

OLS estimates of the effect of preparation on Saber 11° Global Score* with PRE-Saber subsample

Variables	Global score*			Language			Math		
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Preparation	5.4166***	2.3227***	1.3888**	1.2515***	0.5126***	0.3206**	1.0544***	0.6485***	0.4738***
	[0.9600]	[0.6858]	[0.6603]	[0.2116]	[0.1569]	[0.1538]	[0.1860]	[0.1531]	[0.1485]
Global PRE-SABER	0.3436***	0.2465***	0.2432***	0.3517***	0.2429***	0.2392***	0.3751***	0.3009***	0.2977***
	[0.0095]	[0.0079]	[0.0078]	[0.0085]	[0.0071]	[0.0070]	[0.0077]	[0.0070]	[0.0069]
Female		-9.3500***	-9.3619***		-0.3178**	-0.3172**		-3.4565***	-3.4652***
		[0.6191]	[0.5984]		[0.1373]	[0.1329]		[0.1433]	[0.1398]
Group (Strata 1&2 =1)		-13.9872***	-10.9569***		-2.7300***	-2.1455***		-2.3838***	-1.8368***
		[0.8115]	[0.7620]		[0.1690]	[0.1640]		[0.1879]	[0.1770]
Public school			-9.7616***			-1.8330***			-1.7426***
			[1.2650]			[0.2564]			[0.3051]
Constant	173.1014***	279.5472***	290.8963***	38.9465***	58.7205***	60.7340***	39.1227***	61.2333***	63.2778***
	[3.1291]	[7.0830]	[7.0554]	[0.4658]	[1.5325]	[1.5352]	[0.3727]	[1.5116]	[1.5188]
Individual characteristics	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
School characteristics	No	No	Yes	No	No	Yes	No	No	Yes
R-Square	0.333	0.474	0.481	0.205	0.328	0.335	0.32	0.414	0.419
Number of observations	37,263	37,263	37,263	37,263	37,263	37,263	37,263	37,263	37,263

Notes:

Saber Global Score is calculated as a weighted average of the five subjects. Individual characteristics are age, highest level of education of parents (references level is none education) and area of residence (reference level is rural area). School characteristics is meeting time (references level is full-time). Standard errors in square brackets.

* $p<0.1$. ** $p<0.05$. *** $p<0.01$.

Source: Authors' calculations.

Table 10b.
OLS estimates of the effect of preparation on Saber 11° performance with motivation subsample

Variables	Global score			Language			Math		
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Preparation	5.6186***	3.6554***	2.2749***	1.1453***	0.6893***	0.4073***	1.0355***	0.7420***	0.4863***
	[0.5186]	[0.4328]	[0.4250]	[0.1169]	[0.1030]	[0.1022]	[0.1157]	[0.1016]	[0.1005]
Tecnical	-3.8348***	0.3149	-0.3517	-0.5300*	0.2536	0.0899	-0.9882***	-0.1196	-0.2467
	[1.1803]	[1.1265]	[1.1333]	[0.2866]	[0.2729]	[0.2759]	[0.2779]	[0.2670]	[0.2668]
Technological	4.4775***	4.5886***	3.3678***	1.1074***	1.1029***	0.8189***	0.5806**	0.5944**	0.3663
	[1.2512]	[1.1871]	[1.1959]	[0.3016]	[0.2870]	[0.2898]	[0.2951]	[0.2828]	[0.2834]
University	22.4595***	13.1279***	11.0138***	4.9710***	2.9295***	2.4586***	3.8354***	2.2771***	1.8883***
	[1.1180]	[1.0578]	[1.0699]	[0.2643]	[0.2515]	[0.2556]	[0.2571]	[0.2457]	[0.2466]
Female		-8.4900***	-8.3177***		-0.2033**	-0.1745*		-3.3376***	-3.3082***
		[0.4018]	[0.3962]		[0.0931]	[0.0922]		[0.0943]	[0.0935]
Group (Strata 1&2 =1)		-14.1561***	-12.9814***		-2.6251***	-2.4321***		-2.6487***	-2.4671***
		[0.6094]	[0.5795]		[0.1368]	[0.1337]		[0.1406]	[0.1353]
Public school			-4.8320***			-0.7849***			-0.6882***
			[0.6910]			[0.1519]			[0.1570]
Constant	231.6016***	347.0754***	342.9118***	45.9485***	68.5482***	67.2951***	47.0343***	71.5592***	70.7853***
	[1.1041]	[3.6978]	[3.6879]	[0.2635]	[0.9122]	[0.9113]	[0.2558]	[0.9154]	[0.9190]
Individual characteristics	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
School characteristics	No	No	Yes	No	No	Yes	No	No	Yes
R-Square	0.05	0.281	0.295	0.042	0.205	0.215	0.031	0.206	0.214
Number of observations	45,266	45,266	45,266	45,266	45,266	45,266	45,266	45,266	45,266

Saber Global Score is calculated as a weighted average of the five subjects. Individual characteristics are age, highest level of education of parents (references level is none education) and area of residence (reference level is rural area). School characteristics is meeting time (references level is full-time). Standard errors in square brackets
* *p<0.10, ** p<0.05, *** p<0.01*
Source: Authors' calculations.

Table 10c.

OLS estimates of the effect of preparation (accounting for retaking the test) on Saber 11° performance*

Variables	Global score			Language			Math		
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Preparation	7.1892***	4.5973***	3.1042***	1.4864***	0.893***	0.5770***	1.2732***	0.8662***	0.5874***
	[0.3673]	[0.2465]	[0.2331]	[0.0762]	[0.0512]	[0.04947]	[0.0759]	[0.0542]	[0.0516]
Retake		5.6271***	4.9208		0.8659***	0.7207***		-3.3651***	1.0529
		[0.4368]	[0.2331]		[0.8588]	[0.8347]		[0.1009]	[0.9885]
Female		-8.1412***	-8.0206***		-0.05800	-0.04554		-3.3651***	-3.3436***
		[0.2318]	[0.2246]		[0.0470]	[0.0456]		[0.0515]	[0.0504]
Strata 1 & 2 (Yes==1)		-14.3739***	-13.0414***		-2.7710***	-2.5305***		-2.5887***	-2.3594***
		[0.3990]	[0.3596]		[0.0774]	[0.07115]		[0.0860]	[0.0781]
Public school			-5.5101***			-0.9737***			-0.9075***
			[0.5839]			[0.1154]			[0.1223]
Constant	248.7932***	365.0297***	357.2082***	49.8200***	72.6468***	70.8660***	49.9518***	74.775***	73.3080***
	[0.3814]	[1.7447]	[1.7937]	[0.0805]	[0.3816]	[0.3861]	[0.0742]	[0.3960]	[0.4101]
Individual characteristics	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
School characteristics	No	No	Yes	No	No	Yes	No	No	Yes
R-Square	0.005	0.275	0.29	0.004	0.2	0.213	0.003	0.203	0.213
Number of observations	424,589	424,589	424,589	424,589	424,589	424,589	424,589	424,589	424,589

Notes:
*SABER Global Score is calculated as a weighted average of the five subjects. Individual characteristics are age, highest level of education of parents (references level is none education) and area of residence (reference level is rural area). School characteristics is meeting time (references level is full-time). Standard errors in square brackets.
* $p<0.10$, ** $p<0.05$. *** $p<0.01$.
Source: Authors' calculations.

The heterogeneous effects of preparation on students from more-vulnerable socio-economic backgrounds are considerable. Our results demonstrate that students who participate in preparatory activities and are enrolled in public schools earn 5.2 points less (0.12 of a standard deviation) than those enrolled in private schools. Furthermore, low-income students earn 5.8 points less (0.13 of a standard deviation) than middle- and high-income students. These findings suggest that preparatory activities do not narrow the academic achievement gap of students from vulnerable socio-economic backgrounds.

Factors such as motivation and quality of the preparatory activities – which are not directly measured in our complete model– might also be captured in our results. Motivation is a particularly important factor. The slight difference in examination results between girls and boys may be due to both participation in preparatory activities and motivation. Considering that girls are between 20 and 40 percent less likely to participate in preparatory activities, it is possible that those who decide to participate in preparatory activities are more motivated to make good use of the additional materials and preparation, resulting in slightly better results than boys. As part of our research, we included robustness checks to analyze the possible effects of motivation; however, this factor should be further explored in future research.

The quality of the preparatory activities – which is often associated with costs – is another factor that may explain our results. Although the dataset does not allow us to unravel the effects of the quality of the preparatory activities, we were able to verify that insofar as the training is more effective, the chances of enrolling in the activity are considerably reduced for low-income students. The only preparatory activity for which the odds ratios are in favor of low-income students studying at public schools is the activity offered at school with in-house staff. The literature has shown that the education offered at public schools in Colombia is, on average, of lower quality than the education offered at private institutions (Angrist, Bettinger, Bloom, King, & Kremer, 2002), making it reasonable to conclude that the activities offered to the most vulnerable students are of lower quality. In that sense, the participation of low-income students in preparatory activities at their own schools may be perpetuating the achievement gap in the standardized examination. An intervention that should be considered by the local education authorities is to offer professional development to the teachers in public schools who are providing the preparation instruction.

Lowering the costs of high-quality preparation materials may contribute to decreasing the gap caused by unequal access to preparatory activities. As a response to this, the Colombian government has published free study materials on the ICFES web site. They have also distributed additional resources in the national newspaper. However, some students might not be able to obtain this material, because internet access is not universal in Colombia, and the newspaper has a cost. We consider that further actions should be taken. For example, the Government could send study materials directly to the schools to guarantee their availability without

additional costs. The government could also offer differentiated rates, by socio-economic stratum, for the Pre-Saber examination registration fee.

It is important to note that a necessary condition for low-income students to take full advantage of the available materials for preparation is for them to obtain adequate instructions about how to use these materials (Murnane & Willett, 2010). Therefore, in addition to the provision of materials, it is essential to have facilitators who can distribute this material and can explain its use and its scope adequately. In fact, Murnane and Willett (2010) have shown that having close supervision of how students use these materials in their preparatory activities is a key factor in the success of this initiative.

CONCLUSIONS

The returns of preparation, in terms of improving test performance, appear to be modest. We find that, on average, participating in a preparatory activity is associated with a 0.06 SD increase in the global scores on the exit examination. The results indicate that only those students who are enrolled in a Pre-ICFES institution and spend more than three months preparing, experience gains of up to 0.25 SD in their global score.

We also find that those students who benefit less from preparatory activities are public school students and students of low socioeconomic strata. This finding may suggest that preparatory activities can exacerbate the already unequal higher education opportunities. However, we cannot rule out that this finding may have been influenced by unobservable characteristics of the students, such as low motivation. Even though we included robustness checks to test the possible effect of motivation, this factor should be further explored in future research.

There are some limitations of this study that should be acknowledged. Given the non-experimental nature of the data, there is no guarantee that our results can be analyzed as causal estimations of the effect of preparation. However, it is important to highlight that we have taken special actions to reduce the risk of an erroneous estimate of the effect: i) we employed a rich data set, which provides a good set of observable variables and outcomes, ii) we use propensity score matching to properly compare treated and untreated cases, and iii) our results are robust when we take into account some measures that proxy non-observed variables such as ability and motivation.

Our results suggest that schools, parents and teachers should not rely solely on preparatory activities as a sufficient strategy to achieve high scores in exit examinations. It is very unlikely for a student to achieve an excellent score exclusively as a result of a short-term preparation activity (especially if the activity lasts less than 3 months). Other factors can play an essential role in determining individual results. The literature has shown that teacher's efforts, school resources and family characteristics are particularly important in determining success. Students should

obtain a broad and solid academic education during elementary and high school. Obtaining a high score on an exit examination often depends on the knowledge and skills acquired in earlier years. Enrolling in a preparatory activity is probably insufficient to excel if the student does not have a good educational background.

REFERENCES

1. Angrist, J., Bettinger, E., Bloom, E., King, E., & Kremer, M. (2002). Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment. *American Economic Review*, *92*(5). 1535-1558. doi:10.1257/000282802762024629.

2. Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics*, *122*(3). 1235-1264. doi:10.1162/qjec.122.3.1235.

3. Belley, P. & Lochner L. (2007). The Changing Role of Family Income and Ability in Determining Educational Achievement. *Journal of Human Capital*, *1*(1), 37-89. doi:10.1086/524674.

4. Bray, M. (2013). Shadow Education: Comparative Perspectives on the Expansion and Implications of Private Supplementary Tutoring. *Procedia - Social and Behavioral Sciences*, *77*. 412-420. doi:10.1016/j.sbspro.2013.03.096.

5. Bray, M. (2014). The Impact of Shadow Education on Student Academic Achievement: Why the Research is Inconclusive and What Can Be Done about It. *Asia Pacific Education Review*, *15*(3). 381-389. doi:10.1007/s12564-014-9326-9.

6. Briggs, D. C. (2002). *SAT coaching, bias and causal inference* (Doctoral dissertation). Retrieved from https://nepc.colorado.edu/sites/default/files/Briggs_SATCoachingBiasandCasualInference.pdf

7. Buchmann, C., Condron, D. J., & Roscigno, V. J. (2010). Shadow Education, American Style: Test Preparation, the SAT and College Enrollment. *Social Forces*, *89*(2). 435-461. doi:10.1353/sof.2010.0105.

8. Byun, S. Y., & Park, H. (2012). The academic success of East Asian American youth: The role of shadow education. *Sociology of education*, *85*(1), 40-60. doi: 10.1177/0038040711417009

9. Caliendo, M., & Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*, *22*(1), 31-72. doi:10.1111/j.1467-6419.2007.00527.x

10. Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How Does your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR. *Quarterly Journal of Economics*, *126*(4). 1593–1660. doi:10.1093/qje/qjr041.

11. Chica, S. M., Galvis, D. M., & Ramírez, A. (2011). *Determinantes del rendimiento académico en Colombia: pruebas ICFES Saber 11°, 2009* (Documentos de Trabajo Economía y Finanzas No. 11-05). Medellín: EAFIT. Retrieved from https://repository.eafit.edu.co/handle/10784/580

12. Choi, A., Calero, J., & Escardíbul, J. O. (2011). Hell to Touch the SKY? Private Tutoring and Academic Achievement in Korea. In A. Caparrós (Ed.), *Investigaciones de Economía de la Educación*, (Vol. 6., pp. 18-134). Asociación de economía de la educación

13. Dang, H-A. (2007). The Determinants and Impact of Private Tutoring Classes in Vietnam. *Economics of Education Review*, *26*(6). 683-698. doi:10.1016/j.econedurev.2007.10.003.

14. Davis-Kean, P. E. (2005). The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment. *Journal of Family Psychology, 19*(2), 294-304. doi: 10.1037/0893-3200.19.2.294

15. Domingue, B., & Briggs, D. C. (2009). Using linear regression and propensity score matching to estimate the effect of coaching on the SAT. *Multiple Linear Regression Viewpoints*, *35*(1). 12-29.

16. Gamoran, A., & Long, D. A. (2007). Equality of Educational Opportunity A 40 Year Retrospective. In R. Teese, S. Lamb, M. Duru-Bellat, & S. Helme (Eds.), *International Studies in Educational Inequality, Theory and Policy* (pp. 23-47). Springer. doi:10.1007/978-1-4020-5916-2_2

17. Gómez Soler, S. C. (2016). Educational Achievement at Schools: Assessing The Effect of The Civil Conflict Using a Pseudo-Panel of Schools. *International Journal of Educational Development*, *49*(C), 91-106. doi:10.1016/j.ijedudev.2016.02.004.

18. Gómez Soler, S. C. (2017). Civil Conflict and Educational Achievement: The Case of the Colombian Secondary School Exit Examination. *Colombia Internacional*, (92), 73-103. doi:10.7440/colombiaint92.2017.03.

19. Gurun, A., & Millimet, D. L. (2008). *Does Private Tutoring Payoff?* (IZA Discussion Paper No. 3637). Bonn: Institute for the Study of Labor (IZA). Retrieved from http://ftp.iza.org/dp3637.pdf

20. Häkkinen, I., Kirjavainen, T., & Uusitalo, R. (2003). School Resources and Student Achievement Revisited: New Evidence from Panel Data. *Economics of Education Review, 22*(3), 329-335. doi:10.1016/S0272-7757(02)00060-2

21. Hanushek, E. A. (1979). Conceptual and Empirical Issues in the Estimation of Educational Production Functions. *The Journal of Human Resources, 14*(3), 351-388. doi:10.2307/145575

22. Hanushek, E. A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, *24*(3), 1141-1177.

23. Haveman, R., & Wolfe, B. (1995). The Determinants of Children's Attainments: A Review of Methods and Findings. *Journal of Economic Literature, 33*(4), 1829-1878.

24. Heinrich, C., Maffioli, A., & Vazquez, G. (2010). *A Primer for Applying Propensity-Score Matching* (Technical Notes No. IDB-TN-161). Washington: Inter-American Development Bank. Retrieved from https://publications.iadb.org/en/primer-applying-propensity-score-matching

25. Hernandez, O. 2011. *Prácticas en instituciones escolares de la ciudad de Bogotá asociadas con la preparación de la pruebas Saber11* (2011. November). Retrieved from: http://www.icfes.gov.co/documents/20143/234037/Practicas+en+instituciones+escolares+de+bogota+asociadas+a+presentacion+saber+11.pdf

26. Hof, S. (2014). Does Private Tutoring Work? The Effectiveness of Private Tutoring: A Nonparametric Bounds Analysis. *Education Economics*, *22*(4). 347-366. doi:10.1080/09645292.2014.908165.

27. Instituto Colombiano para la Evaluación de la Educación (2010). Orientaciones para el Examen de Estado de la Educación Media (pp. 30). Bogotá: ICFES. Retrieved from: http://www.icfes.gov.co/documents/20143/518161/Guia%20Saber%2011%202010%202.pdf

28. Jayachandran, S. (2014). Incentives to Teach Badly: After-School Tutoring in Developing Countries. *Journal of Development Economics*, *108*. 190-205. doi:10.1016/j.jdeveco.2014.02.008.

29. Krueger, A. B. (2003). Economic Considerations and Class Size. *The Economic Journal*, *113*(485), F34-F63. doi:10.1111/1468-0297.00098

30. Kuan, P-Y. (2011). Effects of Cram Schooling on Mathematics Performance: Evidence from Junior High Students in Taiwan. *Comparative Education Review*, *55*(3). 342-368. doi:10.1086/659142.

31. Liu, J. (2012). Does Cram Schooling Matter? Who Goes to Cram Schools? Evidence from Taiwan. *International Journal of Educational Development*, *32*(1), 46-52. doi:10.1016/j.ijedudev.2011.01.014.

32. Los Mejores Colegios de Colombia según Prueba Saber 11 en 2017. (2017. December). *Revista Dinero*. Retrieved from http://www.dinero.com/edicion-impresa/caratula/articulo/mejores-colegios-de-colombia-segun-prueba-saber-11-en-2017/253328

33. Murnane, R. J., & Willett, J. B. (2010). *Methods Matter: Improving Causal Inference in Educational and Social Science Research*. Oxford, UK: Oxford University Press.

34. Núñez, J., Steiner, R., Cadena, X., & Pardo, R. (2002). ¿Cuáles colegios ofrecen mejor educación en Colombia? (Documento CEDE 2002-03). Retrieved from https://economia.uniandes.edu.co/components/com_booklibrary/ebooks/D2002-03.pdf

35. Ryu, D., & Kang, C. (2013). Do Private Tutoring Expenditures Raise Academic Performance? Evidence from Middle School Students in South Korea. *Asian Economic Journal*, *27*(1)*. 59-83. doi:10.1111/asej.12002.

36. Tobón, D., Posada, H. M., & Ríos, P. (2009). Determinants of the Performance of the Schools in Medellin in the High-School Graduation-Year Test (ICFES). *Cuadernos de Administración*, *22*(38), 311-333.

37. Todd, P. E., & Wolpin, K. I. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal*, *113*(485), F3-F33. doi:10.1111/1468-0297.00097

38. Wößmann, L. (2003). Schooling Resources, Educational Institutions and Student Performance: the International Evidence. *Oxford Bulletin of Economics and Statistics, 65*(2), 117-170. doi:10.1111/1468-0084.00045

39. Zhang, Y. (2013). Does private tutoring improve students' National College Entrance Exam performance?—A Case Study from Jinan, China. *Economics of Education Review*, *32*. 1-28. doi:10.1016/j.econedurev.2012.09.008.

40. Zhang, Y., & Liu, J. (2016). The Effectiveness of Private Tutoring in China with a Focus on Class-Size. *International Journal of Educational Development*, *46*. 35-42. doi:10.1016/j.ijedudev.2015.11.006.