Summer 8-8-2023

Essays on Indirect Effects of Higher Education Policies

Jesús Villero

ABSTRACT

ESSAYS ON INDIRECT EFFECTS OF HIGHER EDUCATION POLICIES

By

JESÚS MANUEL VILLERO AROCA

AUGUST, 2023

Committee Chair: Dr. Daniel Kreisman

Major Department: Economics

This dissertation explores the indirect effects of a large-scale program introduced in Colombia in 2014 that dramatically expanded college financial aid for low-income students. It is composed of two chapters. The first chapter, co-authored with Michael D. Bloem, studies the effect of the increase in post-secondary educational opportunities on teen fertility. Our preferred empirical approach for this chapter uses a triple difference design that leverages variation in the share of female students eligible for the program across municipalities in the country and the fact that the introduction of financial aid should not affect the education and fertility decisions of older women not targeted by the program. We find that fertility rates for women aged 15-19 decreased in more affected municipalities by about 6 percent relative to less affected municipalities. Our results suggest that increasing economic opportunities through expanding college access can contribute to lowering teen fertility rates. The second chapter studies how the dramatic expansion of financial aid for college affected the gender achievement gap in the end-of-high school standardized exam used in post-secondary admission decisions. Using the group of non-low-income students as a comparison group and a non-linear difference-in-differences approach, I provide evidence that the policy caused a reduction in the gender gap at the top by between 9 and 13 percent, starting at the 75th percentile of the distribution of test scores. I present evidence that these results are due to a stronger response among female students (relative to men) and provide a theoretical framework to explain the differential response. The findings have implications for understanding the gender achievement gap in settings with pronounced income-driven barriers to access to higher education.

ESSAYS ON INDIRECT EFFECTS OF HIGHER EDUCATION POLICIES

BY

JESÚS MANUEL VILLERO AROCA

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree
of
Doctor of Philosophy
in the
Andrew Young School of Policy Studies
of
Georgia State University

GEORGIA STATE UNIVERSITY
2023

ACCEPTANCE

This dissertation was prepared under the direction of the candidate's Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Andrew Young School of Policy Studies of Georgia State University.

Dissertation Chair: Dr. Daniel Kreisman

Committee: Dr. Lasse Brune
Dr. Lauren Hoehn-Velasco
Dr. Thomas A. Mroz

Electronic Version Approved:

Ann-Margaret Esnard, Interim Dean
Andrew Young School of Policy Studies
Georgia State University
AUGUST, 2023

Acknowledgments

For Chapter 1, we thank Daniel Kreisman, Jonathan Smith, Lauren Hoehn-Velasco, Thomas A. Mroz, Matias Busso, Juliana Londoño-Vélez, and Lasse Brune for generous feedback and suggestions. We also thank seminar participants at Georgia State University, Universidad del Norte, Universidad Alberto Hurtado, the 2022 Economic Demography Workshop, the 15th Annual Meeting of the Impact Evaluation Network of LACEA, and the 2022 Carolina Region Empirical Economics Day for helpful comments.

For Chapter 2, I am grateful to Daniel Kreisman, Lasse Brune, Lauren Hoehn-Velasco, Thomas A. Mroz, Jonathan Smith, Givi Melkadze, Daniel Millimet, Carolina Concha-Arriagada, Hasan Shahid, Michael Bloem, and seminar participants at Georgia State University for useful feedback.

Table of Contents

List of Tables

List of Figures

Chapter 1: College Opportunity and Teen Fertility: Evidence From *Ser Pilo Paga* in Colombia

(co-authored with Michael D. Bloem)

1.1 Introduction

Teen childbearing is associated with worse outcomes for mothers and their children, including lower educational attainment and poorer labor market outcomes, and with large public costs, including greater reliance on social programs (Kearney and Levine, 2012; Azevedo, Favara, Haddock, López-Calva, Muller and Perova, 2012; Aizer, Devereux and Salvanes, 2022). Not surprisingly, reducing its incidence is a persistent goal among national governments and international agencies. Furthermore, rates of teen childbearing are higher among low-income communities and in places with greater levels of income inequality (Kearney and Levine, 2012, 2014). Youth may be more likely to engage in risky behaviors when chances of economic mobility are low and opportunities to make investments in their own economic progress are limited. Thus, one possible way to break this cycle of early childbearing and poverty is to focus policies on reducing inequality in opportunities for youth to make investments in their own economic progress.

In this paper, we investigate how teen behavior responds to increases in post-secondary educational opportunities by studying the effects on teen fertility of Colombia's 2014 introduction of *Ser Pilo Paga* (roughly translated as "Being a Good Student Pays Off"), a college financial aid program covering full tuition costs at high-quality institutions for high-achieving, low-income students. The Colombian setting is characterized by high teenage fertility rates and high economic inequality, as indicated by the cross-country comparison in Figure A1, with large income-based gaps in college enrollment, high college tuition costs, and little existing access to credit before *Ser Pilo Paga* (SPP).

This setting is suitable for studying this topic because SPP had large educational effects. Londoño-Vélez, Rodríguez and Sánchez (2020a) show that SPP dramatically increased college enrollment on the eligibility margin (57 to 87 percent increases depending on the complier population), virtually eliminating the income-based gap in college enrollment among high-achieving students. Importantly, since colleges increased supply to capture the additional demand, SPP also increased college enrollment among low-income, aid-ineligible students by 14 percent.

There is also evidence that the introduction of SPP altered human capital investment decisions before college. Bernal and Penney (2019) and Laajaj, Moya and Sánchez (2022a) both show that test scores on the national high school exit exam increased among low-income students immediately after the introduction of SPP, particularly at the top of the score distribution. Laajaj et al. (2022a) further show that test scores also increased for low-income students on the national *9th grade* exam, which the authors characterize as a "motivational" effect of SPP. Importantly, they show that these motivational effects on 9th grade test scores reached all the way down to the 19th percentile of the score distribution, illustrating that SPP's behavioral effects extend far beyond the top of the academic distribution when students have more time to prepare to take the high school exit exam that determines their eligibility for SPP.

Given the evidence above, it is reasonable to expect that teens may also alter their behavior on other non-academic dimensions, like decisions about childbearing. We highlight a few notable potential mechanisms for how SPP could affect teenage fertility rates. First, by increasing college attendance, SPP could reduce fertility due to teens having less time to engage in risky behaviors (i.e., a pure incapacitation effect). Second, receiving the scholarship could grant teens greater access to contraceptives through an income effect. Third, SPP can decrease fertility by increasing the opportunity cost of becoming pregnant as a teenager or through a motivational effect that leads students to pursue college attendance opportunities unavailable before SPP. Since the first two mechanisms would largely derive from scholarship recipients after completing high school and enrolling in college, we refer to these potential channels as SPP's "direct" effects on fertility. Conversely, since the third mechanism would mostly derive from students before determining eligibility for SPP, we refer to this channel as SPP's "indirect" (or ex-ante, as in Laajaj et al. (2022a)) effects on fertility.

Our preferred empirical approach uses a triple difference research design leveraging municipality-level variation in SPP eligibility rates determined prior to the introduction of the program and the fact that SPP should not affect the fertility decisions of older women aged 25-29. Eligibility for SPP was based on test scores on the national standardized high school exit exam and scores on a household wealth index. By the time SPP was initially announced, students had already taken the high school exit exam and there was not sufficient time to request a reevaluation of their household wealth index. Our empirical approach uses eligibility rates only from this first cohort of students, who could not influence their scores around the eligibility cutoffs.

We find that fertility rates for women aged 15-19 decreased by about 6 percent in more affected municipalities relative to less affected municipalities. This accounts for approximately one-fourth of the overall decrease in teen fertility observed in the years following SPP's announcement. We rule out that our observed effects are entirely driven by the direct effects of receiving the scholarship upon finishing high school. The timing of the decline in fertility rates—and the fact that the number of fewer births implied by our estimates is larger than the number of actual female SPP scholarship recipients—suggests that incapacitation or income effects of receiving the scholarship itself cannot fully explain the results. We also show results on rates of teen fatherhood, using the more granular data available on father's age that is not available for mother's age, that document effects even for younger teens aged 15-17 who are not old enough to have received the scholarship. Thus, we interpret our findings as largely comprised of indirect effects of SPP, where the new college opportunities created by the program influenced teen fertility decisions before being able to benefit from the program directly. This is consistent with the ex-ante motivational effects on test scores documented by Laajaj et al. (2022a).

We also find that the teen fertility impacts of SPP are larger in municipalities that, before the program, exhibited higher levels of income inequality and a higher share of female students reporting low expectations of enrolling in higher education after finishing high school.[1] These results are broadly consistent with inequality, and "economic hopelessness", being an important determinant of teen childbearing rates (Kearney and Levine, 2014). In addition, we document that the relative reduction in teen births is more prominent in municipalities where female teenagers tend to have children with other teenagers—perhaps indicating a reinforcement of incentives to avoid parenthood when potential fathers also face enhanced college opportunities or increased bargaining power for female teens in such relationships.

Our results are robust to alternative specifications and empirical approaches. We show that the estimated effect of SPP on teen fertility increases in magnitude as a municipality's initial SPP eligibility rate increases, which illustrates that our preferred estimates are not dependent upon how we characterize municipalities as more or less affected by SPP. Furthermore, we see results that are consistent with our main estimates when using a simpler difference-in-differences design, and when using alternative triple difference approaches that rely on different sources of variation. For instance, triple difference results are similar when using a municipality's distance to the nearest SPP-eligible higher education institution instead of variation in SPP eligibility rates, and when using women aged 15-19 with completed education less than eighth grade (and thus likely a school dropout) as a comparison group instead of women aged 25-29. We also rule out that possible confounding events drive our results, including Colombia's peace agreement with the Revolutionary Armed Forces of Colombia, the Zika virus epidemic, and the partial introduction of extended school days in some municipalities.

We add to the literature in three important ways. Our primary contribution is documenting teen fertility responses to a large change in post-secondary schooling *opportunities*, which suggest that improving the future economic prospects of young women through college opportunities can reduce teen pregnancy and early childbearing. An existing literature studies the effects of education on teen pregnancy using exogenous variation from school entry policies and mandatory schooling laws (e.g., Black, Devereux and Salvanes (2008); McCrary and Royer

[1] Figure A2 shows the pre-SPP correlation between adolescent fertility and income inequality and access to higher education in Colombia.

(2011); Alzúa and Velázquez (2017)) and from the duration of school days (Berthelon and Kruger, 2011). Since these policies require additional time to be spent in school, evidence of declines in adolescent fertility in these settings may be due to either an incapacitation effect or a true "human capital" effect of the extra years of contemporaneous education, but not to expanded *future*, *non-contemporaneous* educational opportunities (Doleac and Gibbs, 2016; Alzúa and Velázquez, 2017).

Our analysis represents an empirical test of theoretical predictions that increases in economic opportunities (and increases in opportunity costs) influence the fertility decisions of young women (Becker, 1960; Willis, 1973; Kearney and Levine, 2014). Little is known about how increasing opportunities for schooling affects fertility decisions, where youth still have agency in their schooling choices or where schooling cannot be made compulsory, such as with college attendance. Closest to our work are Cowan (2011) and Koohi (2017) who show that tuition costs at colleges in the United States are positively associated with various risky behaviors of youth, such as the number of sexual partners within the past year (Cowan, 2011) and the prevalence of adolescent childbearing among undocumented Mexican immigrants (Koohi, 2017). We advance this work by exploiting a large-scale program that provides a cleaner shock to post-secondary educational opportunities in a context with more certainty around the labor market returns to investments in higher education and where imperfect credit markets and limited financial aid make it more difficult for low-income students to attend college.

Second, our paper is related to a literature that studies the teen fertility impacts of interventions in developing countries that aim to improve economic opportunities and empowerment for adolescent women (Jensen, 2012; Duflo, Dupas and Kremer, 2015, 2021; Muralidharan and Prakash, 2017; Bandiera, Buehren, Burgess, Goldstein, Gulesci, Rasul and Sulaiman, 2020). We extend this body of work by providing evidence on how opportunities for college attendance, rather than primary or secondary schooling, relates to adolescent fertility decisions. This evidence is particularly relevant for countries where, like Colombia, secondary schooling is relatively accessible and where attending college is increasingly important for economic mobility. Third, by examining understudied non-educational outcomes (Cowan, 2011; Doleac and Gibbs, 2016; Koohi, 2017), we build on the literature of the effects of the *Ser Pilo Paga* program (Londoño-Vélez et al., 2020a; Bernal and Penney, 2019; Laajaj et al., 2022a) and the effects of college financial aid programs more broadly on the decisions of high school students (Cáceres-Delpiano, Giolito and Castillo, 2018).

The remainder of this paper is organized as follows: In the next section, we discuss the Colombian context and describe the details of the *Ser Pilo Paga* program. The third section describes the data sources we use, discusses the key variables used in our analyses, and presents trends in fertility rates in Colombia. The fourth section discusses our identification strategies and estimation approaches. The fifth section presents our core empirical results and section six tests for the sensitivity and robustness of those results. Finally, section seven concludes.

1.2 Background

1.2.1 Teen Fertility in Colombia

Similar to many Latin American and Caribbean countries, teen fertility is high in Colombia. Estimated at 70.7 births per 1,000 women aged 15-19 years in 2014 (when SPP was announced), the adolescent fertility rate in Colombia was slightly higher than the Latin American average, more than twice that of other countries with similar income levels and nearly three times higher than in the United States.[2] These "higher-than-expected" adolescent fertility rates observed in Latin American countries are likely associated with the high levels of inequality of income (and opportunities) observed in the region (Azevedo et al., 2012).

In contrast, in 2014, Colombia had a lower *total* fertility rate than the average Latin American country, similar to the overall fertility rates in other upper middle-income countries and the United States. However, about 22% of the overall number of births in the country were from

[2] As a region, Latin America and the Caribbean has the second highest fertility rate for teenagers globally, second only to Sub-Saharan Africa. A general discussion about this phenomenon can be found in Azevedo et al. (2012). A cross-country comparison, highlighting Colombia, is presented in Figure A1. Data are from the World Bank's World Development Indicators.

mothers aged 19 or younger that year. As a result, early childbearing is a worrisome phenomenon and a policy concern in Colombia, given its association with worse prospects for the adolescent mothers and their children in terms of health, education, and labor market outcomes (Gaviria, 2010; Azevedo et al., 2012; Urdinola and Ospino, 2015).

Early parenting in Colombia is primarily a female phenomenon. Data from the most recent Demographic and Health Survey (2015) show that adolescent women are 6.4 times more likely to have at least one child than adolescent men—13.6 percent versus 2.1 percent (Flórez and Soto, 2019). Furthermore, birth records data indicate that only 22 percent of births to adolescent women between 2008 and 2014 had a teenage father.[3] While teenage pregnancy affects all income groups, it is particularly worrying among low-income women. Low-income Colombian teenagers are five times more likely to have ever been pregnant than their high-income peers (Flórez and Soto, 2019).

In the last decade and a half, Colombia has implemented several programs and policies directly aimed at reducing teenage pregnancies.[4] Among the most relevant initiatives is the implementation of the Youth Friendly Health Services Model (SSAAJ, from the Spanish acronym) and the Program of Education in Sexuality and Construction of Citizenship (PESCC), both launched in 2007-2008 and scaled up nationally in subsequent years.[5] In 2012, the national government additionally launched a strategic framework to address the issue comprehensively, articulating different actors within the public sector.[6] On top of others not directly targeted at reducing fertility like *Familias en Acción*, the conditional cash transfer program in Colombia, these initiatives likely contributed to the downward trend in teenage fertility observed in the country since the mid-2000s after a concerning period of increase during the 1990s (Flórez and Soto, 2019; Attanasio, Sosa, Medina, Meghir and Posso-Suárez, 2021).[7]

[3] The age of consent in Colombia is 14 years old.

[4] See part three of Vargas Trujillo, Flórez, Cortés and Ibarra, eds (2019) for a recent review.

[5] *Modelo de Servicios de Salud Amigables para Adolescentes y Jóvenes* (SSAAJ) and *Programa de Educación para la Sexualidad y Construcción de Ciudadanía* (PESCC) in Spanish.

[6] National Department of Planning (DNP). *Documento CONPES Social* No. 147.

[7] Since all these policies were implemented years before SPP was introduced, we do not view them as threats to our identification strategy, but rather as possible factors explaining the decline in adolescent fertility observed before SPP.

1.2.2 Ser Pilo Paga and Higher Education in Colombia

Ser Pilo Paga was announced by surprise on October 1st of 2014 by President Santos's administration. The program was publicly funded and covered recipients' full tuition cost of attending an undergraduate program at any university in Colombia with a High Quality Accreditation. The aid came in the form of a loan that is forgiven upon graduation, although only about 1.9 percent of SPP beneficiaries from the first three cohorts had dropped out of the program (Londoño-Vélez et al., 2020a).[8] Additionally, SPP recipients would receive a biannual stipend of at least the national minimum wage to help cover students' living expenses.

Eligibility for SPP was based on both need and merit. First, students must score above a cutoff on the SABER 11, which is similar to the SAT in the United States. The SABER 11 exam is taken by nearly all high school seniors regardless of their plans to attend an institution of higher education. SABER 11 scores play a significant role in college admissions, with about four-fifths of institutions using them in admissions considerations (OECD and World Bank, 2012). The SABER 11 cutoff score was placed at approximately the 91st percentile for the first cohort of SPP.

Second, students must be below a cutoff on the SISBEN, Colombia's wealth index used to target social welfare programs. The SISBEN cutoff varies by geographic location. The cutoff is 57.21 (over 100) in the 14 main metropolitan areas, 56.32 in other urban areas, and 40.75 in rural areas. Between 2015 and 2018, there were about 10,000 SPP beneficiaries per year (43% of them women), which represents about one-third of students attending an institution with High Quality Accreditation.

In the first year of the program, students had already taken the SABER 11 exam before SPP was announced. Moreover, there was insufficient time to request a reevaluation of their household wealth index before determining eligibility for SPP. Thus, students in this first cohort had no opportunities to influence their test scores or wealth index scores in response to the SPP eligibility cutoffs.

[8] The SPP program considers students to have dropped out if they have not attended a high-quality institution for three or more consecutive semesters.

Tuition at the high-quality private universities is very expensive, both compared to private universities in other countries and to the public universities in Colombia (OECD and World Bank, 2012). Since the tuition at the high-quality public universities is relatively low, these institutions are historically oversubscribed, leading to highly selective admissions. Prior to SPP, there were very few financial aid opportunities for high-achieving, low-income students. Only 11 percent of first-year undergraduate students had a student loan before SPP (Ferreyra, Avitabile, Botero Álvarez, Haimovich Paz and Urzúa, 2017).

1.3 Data and Key Variables

This section describes our data sources and key variables. We gather data from publicly available sources on births and population counts in Colombia in order to calculate age-specific fertility rates. To compute a measure that indicates which municipalities were more or less affected by SPP, we collect SABER 11 test score data to calculate SPP eligibility rates.

1.3.1 Data Sources

We use the universe of birth records and annual population estimates from the Colombian National Department of Statistics (DANE, from the Spanish acronym) from 2008 to 2020. Individual birth records contain information about the mother's age in 5-year intervals (i.e., 15-19, 20-24, 25-29, etc.) and about her municipality of *residence* (in addition to where the birth took place). The records also contain the year and month of occurrence of each birth. We use these data to create a municipality by age group and year panel dataset of age-specific fertility rates, which is our primary outcome.

We also use administrative data from the Colombian Institute for the Assessment of Education (ICFES) containing student-level information of the national standardized high school exit exam, SABER 11, including test scores and socio-demographic characteristics. Importantly, these data include information about SISBEN eligibility and the municipality of residence of the student.

Finally, we complement our data with pre-SPP municipality characteristics which we obtain from the Center for the Study of Economic Development (CEDE) from Universidad de los Andes, the Ministry of Education, and DANE.

1.3.2 Construction of Analysis Measures

We use the birth records and population estimates to create a municipality-of-residence by age group panel dataset of age-specific fertility rates, our primary outcome. Our main estimates use the natural log of these fertility rates. Throughout the descriptive and econometric analyses that follow, we account for the lag between conception and birth by using the year-month of birth and the gestational age at delivery to approximate the year-month of conception of each newborn. For 80 percent of births in our sample, this is equivalent to assuming that conception occurred nine months before the reported date of birth.

We define our municipality-level treatment intensity measure as the rate of female SABER 11 test takers in 2014 who are eligible for SPP in each municipality.[9] We then separate the sample at the median, the top half representing the *treatment municipalities* and the bottom half representing the *comparison municipalities*. We do not observe the exact SISBEN score of the students in the SABER 11 data and, therefore, their precise eligibility on the SISBEN margin. However, students report if they are categorized as SISBEN level 1 or 2. A SISBEN level of 1 or 2 is roughly equivalent to being eligible for SPP on the SISBEN margin, whereas students with higher SISBEN levels or not categorized are typically ineligible. On the SABER 11 margin, we determine students' eligibility using their test scores and the SPP threshold established by the government for 2014. For these students, the SPP program was announced after they had taken the SABER 11 exam. Thus, our eligibility rates avoid possible endogenous responses to the announcement of the program or its eligibility thresholds.

[9] ICFES administers the SABER 11 exam in both the spring and fall semesters each year, with the vast majority of students taking the exam in the fall semester. SPP eligibility on the SABER 11 margin was based on exams taken in the fall semester. Typically, only students in a limited set of private schools whose academic calendar is synchronized with the United States take the SABER 11 exam during the first (spring) semester of the year. For example, in 2014 (the year when SPP was introduced), 95.6 percent of the test takers took the test in the second (fall) semester.

We attempt to assess the validity of our treatment intensity measure by estimating whether it is associated with an increase in SABER 11 test scores after SPP is introduced. This is essentially testing whether we can replicate the results from Bernal and Penney (2019) and Laajaj et al. (2022a) using our treatment measure. We use individual-level data on female SABER 11 test takers between 2010 and 2016 and estimate a triple difference model that compares standardized test scores of SISBEN-eligible students between treatment and comparison municipalities. See Appendix B. for a full description of this analysis. Consistent with the existing evidence, we find that, after the introduction of SPP, SABER 11 test scores increased in treatment municipalities for SISBEN-eligible students by about 0.03 standard deviations, relative to comparison municipalities. These findings support the notion that our treatment intensity measure is adequately capturing the mechanisms underlying the introduction of SPP.

1.3.3 Analytic Sample and Summary Statistics

We restrict our sample to municipalities (i) with at least one observed birth from each age group we use in our empirical analysis for all years between 2008 to 2020 and (ii) with SABER 11 information in 2014. By doing this, we drop extremely small municipalities from our sample. Our results are not sensitive to the exclusion of these municipalities. Our final analysis sample consists of a balanced panel of 1,067 municipalities for conception years 2008-2019 (out of 1,122 in the country and 1,105 with SABER 11 information in 2014).

Table 1 displays means and standard deviations of SPP eligibility rates in 2014, weighted by the number of female students in each municipality, for both treatment and comparison municipalities. Comparison municipalities had about 20 fewer SPP eligible students per 1,000 female students in 2014. Figure A3 plots the full distribution of SPP eligibility rates for the municipalities in our sample. About 36 percent of municipalities had zero SPP eligible female students in 2014. We do not interpret these municipalities as completely "untreated." Since, as Londoño-Vélez et al. (2020a) and Laajaj et al. (2022a) document, SPP had effects on students throughout the distribution of students' achievement, our view is that students did not need to be eligible for SPP to be affected by the introduction of the program. We use eligibility rates (in 2014) to characterize municipalities as more or less affected by SPP.

Table 1. Summary Statistics of Key Variables

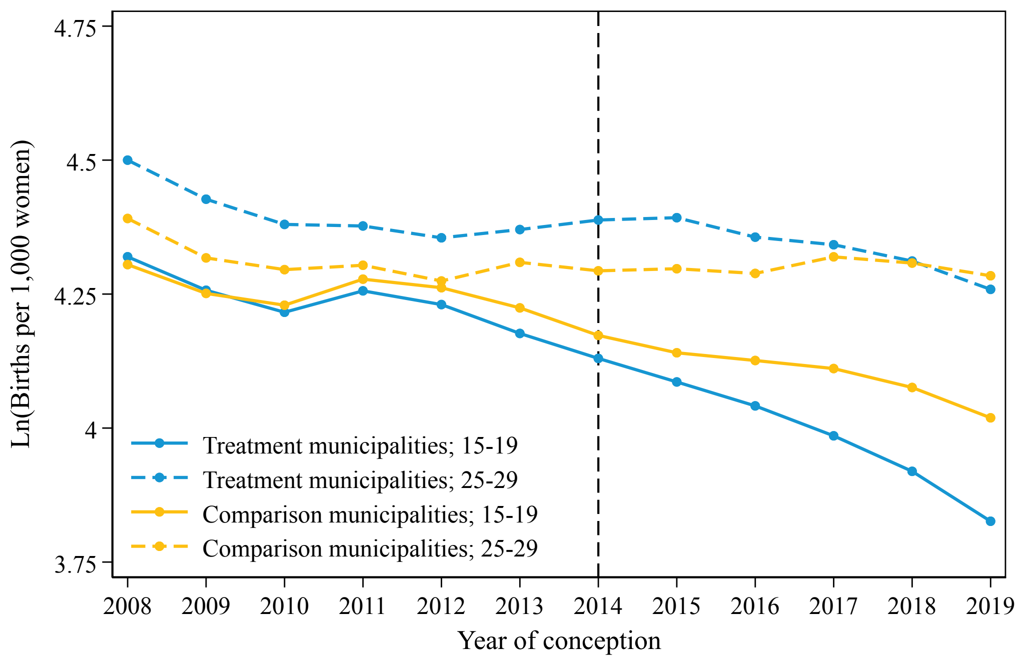
	Treatment municipalities	Comparison municipalities	All
Pre-SPP births per 1,000 women (2008-2013)			
Age 15-19	72.1	77.1	73.6
	(20.2)	(29.6)	(23.5)
Age 25-29	84.0	80.2	83.0
	(20.8)	(27.7)	(23.0)
SPP eligibility rates (2014)			
Per 1,000 female students	26.8	6.6	21.6
	(15.5)	(4.8)	(16.2)
Number of municipalities	541	526	1,067

Notes: This table shows means and standard deviations (in parentheses) for age-specific municipality-level birth rates and SPP eligibility rates between treatment and comparison municipalities. Birth rates are averages from 2008 to 2013 and are weighted using each municipality's annual age-specific female population. SPP eligibility rates are from 2014, the first cohort of students eligible for SPP, and are averaged using the number of female students in each municipality as weights. Treatment municipalities are above the median in female eligibility rates for SPP in 2014, while comparison municipalities are below the median.

Figure A4 visualizes the municipality-level variation in the discrete version of SPP eligibility rates in 2014. While there are some clusters of treatment municipalities at a local level, there are treatment and comparison municipalities in every region of Colombia. Table A1 suggests that these two groups of municipalities were different, on average, in terms of pre-SPP characteristics. For example, treatment municipalities have larger populations, lower poverty levels, and higher secondary school enrollment rates. Importantly, since our identification strategy relies on an assumption of parallel fertility rate trends in absence of SPP, these differences do not invalidate our empirical strategy. Figure A4 also shows the municipalities that had at least one SPP-eligible higher education institution. There were 15 municipalities with an SPP-eligible institution in 2014. This increased to 20 in 2016 and 21 in 2017.

Table 1 also displays means of municipality-level fertility rates during the pre-SPP period for both 15-19 year olds and 25-29 year olds, weighted by the annual age-specific female population. These are the two age groups we use in our triple difference empirical strategy, which we describe in detail in the next section. Compared to the comparison municipalities, treatment municipalities have slightly lower fertility rates for women aged 15-19, but slightly higher fertility rates for women aged 25-29. Figure A5 plots the complete distribution of adolescent fertility rates, which shows a substantial amount of overlap between the distributions of treatment and comparison municipalities. Finally, Figure 1 shows the raw trends in average log fertility rates between age groups for both treatment and comparison municipalities. This figure mimics our empirical approaches discussed in the next section.

Figure 1. Trends in Fertility Rates for 15-19 Year Olds and 25-29 Year Olds by Treatment and Comparison Municipalities



Notes: This figure plots trends in average log fertility rates between age groups for both treatment and comparison municipalities. Treatment municipalities are above the median in female eligibility rates for SPP in 2014, while comparison municipalities are below the median. The averages weight municipalities by the annual population of women in each municipality and age group.

1.4 Empirical Analysis

To estimate the effect of SPP on teen fertility, we follow difference-in-differences approaches. Our designs exploit variation in the share of female students eligible for the program across municipalities. Our preferred triple difference approach also leverages the fact that the introduction of SPP did not affect the education and fertility decisions of older women not targeted by the program. This section first describes the identifying assumptions behind our research designs, and then presents the estimation procedures.

1.4.1 Identification Strategies

We use two approaches to estimate the impact of SPP on the adolescent fertility rate. First, we estimate a simple difference-in-differences model that compares the fertility rate of 15-19 year olds before and after the introduction of SPP between municipalities with different eligibility rates for the program. The identifying assumption underlying this approach is that the trends in birth rates observed in comparison municipalities provide a good counterfactual for the trends in treatment municipalities in the absence of SPP.

Our second and preferred approach is a triple difference model that additionally uses women aged 25-29 as a within-municipality comparison group. We choose 25-29 year olds as our within-municipality comparison group because it is the group closest in age to the 15-19 year olds that is likely not affected by the introduction of SPP. The 25-29 year old group cannot be SPP beneficiaries and most are likely past their college-going years. The 20-24 group is partially affected by the introduction of SPP during our sample period given the nature of our birth records data and is also more likely to be affected by the general equilibrium effects of SPP on the higher education market.

The identifying assumption for this triple difference design is that in the absence of the policy, the differentials in fertility outcomes between 15-19 and 25-29 years old in municipalities with higher SPP eligibility rates (treatment municipalities) would have evolved similarly to these differentials in municipalities with lower SPP eligibility rates (comparison municipalities). This is the usual parallel trends assumption underlying double difference-in-differences designs applied to the triple difference case. We provide evidence in support of this assumption when we discuss our main results in subsection 1.5.

We prefer the triple difference approach because it better mitigates potential bias coming from unobserved, *time-varying* heterogeneity across municipalities. By including a within-municipality comparison group, the triple difference approach accounts for municipality-specific factors that might coincide with the introduction of SPP that the simple difference-in-differences approach cannot account for. While we provide empirical support that the parallel trends assumptions are satisfied in both the difference-in-differences and triple difference approaches, we view the identifying assumptions in the triple difference approach as more theoretically plausible.

The main identification threat to the preferred triple difference strategy is the existence of other confounding events or policies that could have differentially affected the fertility rate of women in different age groups *and* are also correlated with our municipality-level treatment variable. We provide evidence that other majors events that occurred in the country around 2014 cannot explain our results in subsection 1.6.

1.4.2 Estimation Approaches

We implement the difference-in-differences design with an event-study specification that we estimate by Ordinary Least Squares (OLS) using the sample of 15-19 year olds. More specifically, we use the following specification:

$$Y_{mt} = SPP^*_m \times \sum_{\substack{\tau=-7 \\ \tau\neq-1}}^{4} \alpha_\tau \mathbb{1}[t = \tau] + \theta_m + \theta_{d(m)t} + \epsilon_{mt}, \tag{1}$$

where Y_{mt} is the log of the teen fertility rate in municipality m and year $t \in [-7, 4]$, which is measured in years relative to October 1, 2014, the date in which SPP was first announced. In Equation 1, SPP^*_m denotes treatment and comparison municipalities and is defined as

$SPP^*_m = \mathbb{1}\left[SPP_m > \text{median}\left(SPP_m\right)\right]$ with SPP_m being the rate of female students eligible for the program in a given municipality in 2014.[10] θ_m and $\theta_{d(m)t}$ are municipality fixed effects and year fixed effects that we allow to be department-specific, respectively.[11] Finally, ϵ_{mt} is an error term.

In a similar fashion, we implement our preferred triple difference design with the following specification that we also estimate by OLS but using the sample of 15-19 and 25-29 year olds:

$$Y_{amt} = Teen_a \times SPP^*_m \times \sum_{\substack{\tau=-7 \\ \tau\neq-1}}^{4} \beta_\tau \mathbb{1}[t=\tau] + \gamma_{am} + \gamma_{mt} + \gamma_{ad(m)t} + \varepsilon_{amt}, \tag{2}$$

where Y_{amt} is the log fertility rate of age group $a \in \{15\text{-}19, 25\text{-}29\}$ in municipality m and year $t \in [-7, 4]$, again measured in years relative to the announcement of SPP. $Teen_a$ is an indicator for age group defined as $Teen_a = \mathbb{1}\left[a = 15\text{-}19\right]$. γ_{am} and γ_{mt} are age group by municipality and municipality by year fixed effects, respectively. $\gamma_{ad(m)t}$ are age group by year fixed effects that we also allow to be department-specific. Finally, ε_{amt} is an error term.

In Equation 1 and Equation 2, $\alpha_{\tau,t\geq0}$ and $\beta_{\tau,t\geq0}$ represent the average treatment effect of SPP on teen fertility at time $t = \tau$ after the introduction of the program. For the estimation, we use the year before the announcement of SPP ($t = -1$) as our reference period. For Equation 2, and in the standard way for triple difference specifications, we include three two-way interactions between age groups, municipalities, and years. The age group by municipality fixed effects (γ_{am}) control for time-invariant, municipality-specific factors (both observed and unobserved) that affect fertility rates and that are potentially different by age groups. The municipality by year fixed effects (γ_{mt}) control for municipality-specific trends in fertility rates common to all age groups. Finally, the age group by department year effects $\left(\gamma_{ad(m)t}\right)$ account for age-specific

[10] We show in subsection 1.6 that our results are robust to different estimation approaches, including using a continuous version of our treatment intensity measure, excluding capital cities from the sample, using all available municipalities with the inverse hyperbolic sine transformation of fertility rates as the outcome, and using fertility rates in levels as the outcome estimated by either OLS or Poisson models.

[11] Departments in Colombia are similar to states in the United States. A group of municipalities forms each department.

trends in fertility and arbitrary shocks to fertility that are common to all municipalities in a given region. As mentioned earlier, the remaining and identifying source of variation we leverage is the *differential* effect that SPP had on the adolescent fertility rate in the treatment municipalities (relative to the comparison municipalities in the same department).

To summarize the event-study estimates of SPP's effects in a single estimate, we also estimate a version of Equation 1 and Equation 2 that replaces the year indicators with a single post-SPP indicator variable. Specifically, we estimate the following two equations by OLS:

$$Y_{mt} = \alpha \left(SPP_m^* \times Post_t\right) + \theta_m + \theta_{d(m)t} + \xi_{mt}, \text{ and} \tag{3}$$

$$Y_{amt} = \beta \left(Teen_a \times SPP_m^* \times Post_t\right) + \gamma_{am} + \gamma_{mt} + \gamma_{ad(m)t} + u_{amt}, \tag{4}$$

where $Post_t = \mathbb{1}[t \geq 0]$ and everything else is defined as in Equation 1 and Equation 2. In Equation 3 and Equation 4, α and β are the summary difference-in-differences and triple difference parameters across all post-SPP years, respectively.

In all our regressions, we cluster the standard errors at the municipality level and weight each cell by the population of women in each municipality and age group. Excluding weights would give each municipality equal weight in the regression by default, which we argue is inappropriate in our context, particularly due to the large population variation across municipalities in Colombia.[12]

1.5 Results

This section reports and discusses our results. We begin by presenting our difference-in-differences estimates of the teen fertility impacts of *Ser Pilo Paga*. We then present our preferred estimates which use a triple difference approach. Finally, we present supplemental analyses to explore the mechanisms and heterogeneity of the main results.

[12] If we were able to run an individual-level regression, there would naturally be many more observations from municipalities with larger populations. Our use of population weights essentially approximates the use of individual-level data.

1.5.1 Difference-in-Differences Estimates

Figure 2 shows the event study estimates of α_τ from the simple difference-in-differences specification in Equation 1 that compares adolescent fertility rates across municipalities with higher and lower initial SPP eligibility rates, before and after the introduction of the program. None of the coefficients in the pre-period are statistically significant at the 5 percent level. Following Borusyak, Jaravel and Spiess (2021), we conduct a more formal test of pre-period trends by estimating the difference-in-differences specification using only the set of untreated observations and running a joint F-test of these coefficients. With a p-value of 0.711, this test cannot reject the null hypothesis that the pre-period coefficients are jointly equal to zero.

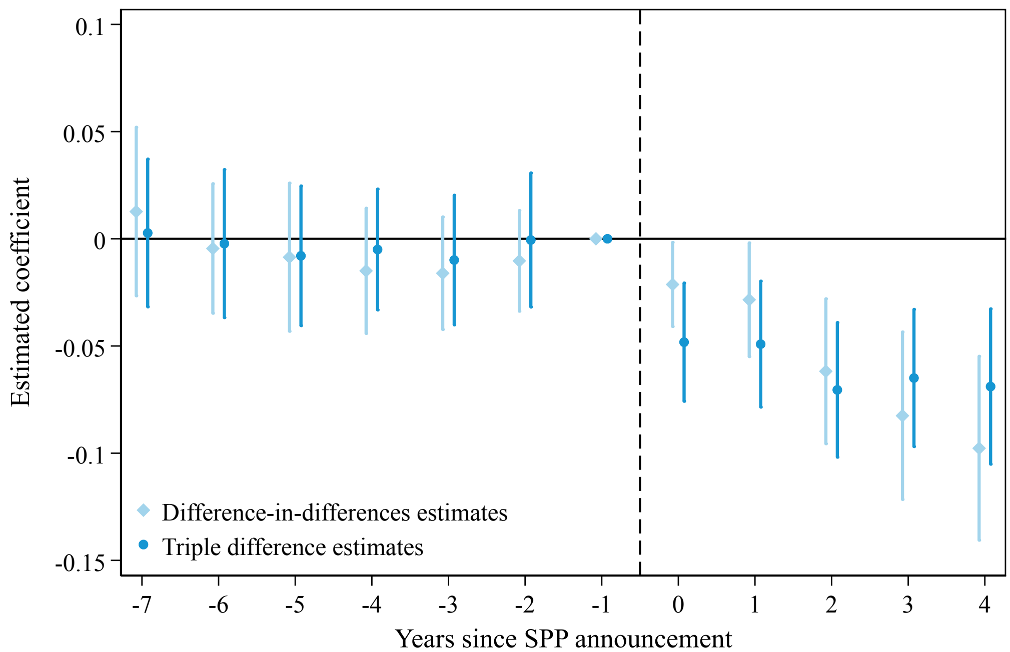
After the introduction of SPP, fertility rate trends between treatment and comparison municipalities change significantly, with teen fertility rates decreasing more in treatment municipalities. All post-period coefficients are negative and statistically significant at the 5 percent level. Column 1 of Table 2 presents the summary estimate from the difference-in-differences design which implies that after SPP was introduced fertility rates of women aged 15-19 in treatment municipalities decreased by about 5.2 percent relative to comparison municipalities.[13]

In support of our use of initial SPP eligibility rates to characterize municipalities as more or less affected by SPP, we show that the teen fertility impacts of SPP are larger in municipalities with higher eligibility rates. To do this, we replace the indicator for being above the median in SPP eligibility with indicators for the *quartile* of SPP eligibility rates. Municipalities in the first (lowest) quartile of eligibility become the reference group in the regression, which we note all had zero female students eligible for SPP in 2014. Column 3 of Table 2 reports the summary results, but we also show the event-study equivalent of this specification in Figure A7, Panel (a).

Relative to municipalities in the 1st quartile of SPP eligibility, we estimate that 2nd quartile municipalities experienced a 3.1 percent decrease (not statistically significant) in adolescent fertility. Meanwhile, we estimate that the 3rd and 4th quartile municipalities

[13] The exact percentage changes implied by our estimates are given by $100 \times [\exp(\hat{\alpha}) - 1]$.

Figure 2. Event Study Estimates of the Effect of *Ser Pilo Paga* on Teen Fertility Rates



Notes: This figure plots the difference-in-differences estimates of α_τ from Equation 1 and the triple difference event study estimates of β_τ from Equation 2. The dots and diamonds represent the estimated coefficients and the vertical lines represent 95 percent confidence intervals. All estimates are weighted by the annual population of women in each municipality and age group. Standard errors are clustered at the municipality level.

experienced a 6.9 and 7.9 percent decrease, respectively. The estimates for the 3rd and 4th quartile municipalities are both statistically different from the estimate for the 2nd quartile at the 10 percent significance level. These results highlight that our estimates are not dependent upon how we split municipalities into treatment and comparison groups.

1.5.2 Triple Difference Estimates

Figure 2 also displays our preferred estimates of the fertility impacts of SPP (β_τ) using the triple difference event-study specification in Equation 2. Similar to the difference-in-differences results, the pre-period coefficients are again close to zero and not statistically significant. An F-test of the null hypothesis that the pre-period coefficients are jointly zero using only the set of untreated observations produces a p-value of 0.985 (Borusyak et al., 2021). This provides

Table 2. Summary Difference-in-Differences and Triple Difference Estimates of the Effect of *Ser Pilo Paga* on Teen Fertility Rates

	Log fertility rate			
	(1)	(2)	(3)	(4)
$SPP \times Post$	-0.052***			
	(0.016)			
$Teen \times SPP \times Post$		-0.057***		
		(0.011)		
$SPPQuartile \times Post$				
1st quartile			[Reference]	
2nd quartile			-0.031	
			(0.027)	
3rd quartile			-0.069***	
			(0.018)	
4th quartile			-0.079***	
			(0.021)	
$Teen \times SPPQuartile \times Post$				
1st quartile				[Reference]
2nd quartile				-0.040**
				(0.019)
3rd quartile				-0.079***
				(0.015)
4th quartile				-0.091***
				(0.018)
Observations	12,804	25,608	12,804	25,608
Treatment municipalities	541	541	–	–
Comparison municipalities	526	526	–	–
Pre-trends test p-value	0.711	0.985	–	–

Notes: This table presents summary difference-in-differences and triple difference estimates using Equation 3 and Equation 4, respectively. Column 1 presents the difference-in-differences estimates. Column 2 presents the main triple difference estimates. Columns 3 and 4 use indicators for the quartile of SPP eligibility rates instead of an indicator for being above the median in SPP eligibility. The reference group here are municipalities in the first (lowest) quartile of SPP eligibility. All estimates are weighted by the annual population of women in each municipality and age group. Standard errors are clustered at the municipality level and presented in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

empirical support in favor of the parallel trends assumption of the triple difference research design.

Starting in the first year after the introduction of SPP, there is a distinct decrease in the fertility rate of women aged 15-19 in treatment municipalities relative to comparison municipalities. All post-SPP coefficients are negative and statistically significant at the 5 percent level. We report the summary estimates using Equation 4 in column 2 of Table 2. The triple difference estimate in this specification indicates that SPP reduced fertility rates of women aged 15-19 in treatment municipalities by 5.7 percent relative to comparison municipalities.[14] This effect accounts for approximately one-fourth of the overall decrease in adolescent fertility observed in Colombia in the years following the announcement of SPP.

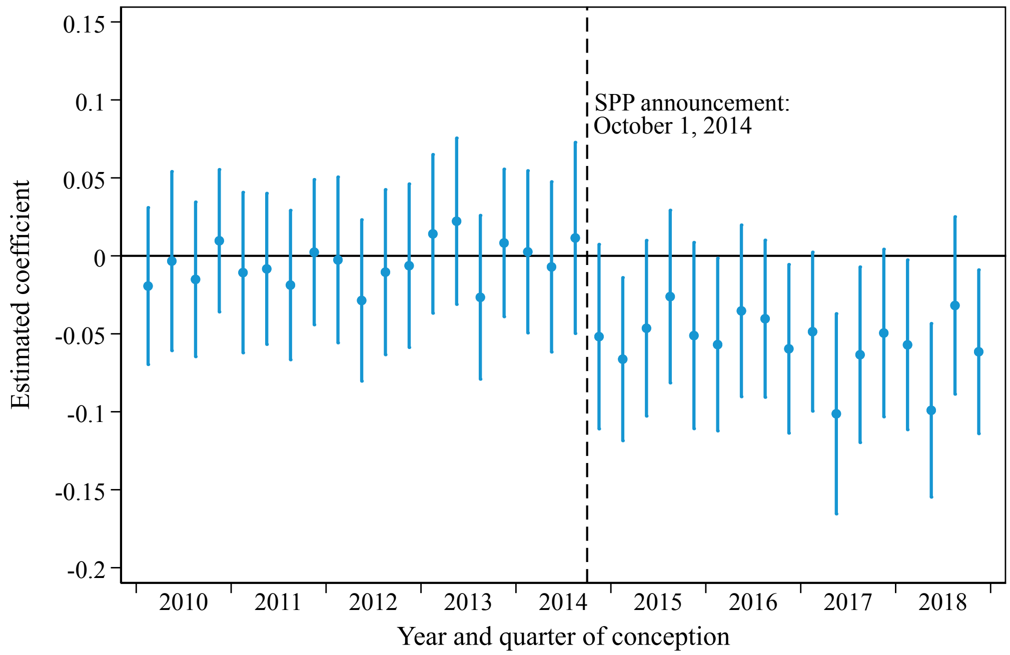
To illustrate even more clearly that the timing of the changes in fertility rate trends aligns with the timing of the introduction of SPP, we estimate Equation 2 using a quarterly-level dataset.[15] SPP was announced on October 1st in 2014. Thus, for SPP to be responsible for the relative decrease in adolescent fertility that we observe, we would expect quarterly-level effects of SPP to be apparent starting exactly in the fourth quarter of 2014. Indeed, this is what we observe from the quarterly-level estimates in Figure 3. Moreover, the summary estimate at the quarterly level is identical to our main estimate at the annual level.

Next, we show that the estimated effect of SPP increases (in magnitude) as a municipality's initial SPP eligibility rate increases. To show this, we again use indicators for the *quartile* of SPP eligibility rates instead of an indicator for being above the median in SPP eligibility. The reference group now becomes municipalities in the first (lowest) quartile of SPP eligibility. The summary results are reported in Figure A6 and in Table 2. The estimates from the event-study equivalent of this specification are shown in Figure A7, Panel (b). Relative to municipalities in the 1st quartile of eligibility, 2nd quartile municipalities experienced a 4 percent decrease in fertility rates, while the 3rd and 4th quartile municipalities experienced a 7.9 and 9.1 percent decrease, respectively. The estimates for each quartile are statistically different from the

[14]The exact percentage changes implied by our estimates are, again, given by $100 \times [\exp(\hat{\beta}) - 1]$.

[15]Since there are some municipalities that have zero births in some cells at the quarterly level, we use the inverse hyperbolic sine transformation of the birth rate as the outcome instead of the log transformation. The interpretation of the estimated coefficients from this model is similar to our main log model. See Bellemare and Wichman (2020) for a discussion and formal derivation. In Figure A8, we show our quarterly results hold using a Poisson model.

Figure 3. Triple Difference Event Study Estimates of the Effect of *Ser Pilo Paga* on Teen Fertility Rates Using Quarterly Data



Notes: This figure plots the triple difference event study estimates of β_τ from Equation 2 using quarterly data instead of annual data. In contrast to the annual specification we use the inverse hyberbolic sine transformation of the birth rate as the outcome instead of the log transformation, because at the quarterly level, some municipalities have zero births in some of the cells. The quarters in 2008 are the reference period. Only estimates for four years around 2014 are plotted. The dots represent the estimated coefficients and the vertical lines represent 95 percent confidence intervals. All estimates are weighted by the annual population of women in each municipality and age group. Standard errors are clustered at the municipality level.

1st quartile at the 5 percent level (1 percent for the 3rd and 4th quartiles), and the estimates for the 3rd and 4th quartiles are statistically different from the estimate for the 2nd quartile at the 1 percent level.

These results highlight two important points. First, our main estimates are not purely the product of a fortuitous split of municipalities at the median of SPP eligibility. Second, our estimates represent relative effects, not total effects. We use the initial SPP eligibility rate to compare municipalities that are plausibly more and less affected by this nationally implemented program. That the estimated effects for municipalities in the 3rd and 4th quartile of eligibility are larger in magnitude than the main estimate suggests that the *total* effects of SPP on adolescent fertility rates in Colombia are likely also larger than what our main estimates indicate.

1.5.3 Mechanisms

In this section, we aim to assess the extent to which our estimates are explained by direct effects—where receipt of the scholarship itself drive the results through incapacitation (via college attendance) or income effects—or by indirect effects—where results are driven by behavioral responses before being able to receive the scholarship, such as motivational effects, changing opportunity costs, or peer effects. Since our data limit us to estimate effects on women within an age range of 15 to 19 years old, our main estimates may only reflect the direct effects of students receiving SPP, going to college, and reducing (or delaying) childbearing that would have occurred during their teen college-age years (i.e., 18 or 19 years old).

To analyze these mechanisms, we first assess the extent to which the direct effects explain our results. To do this, we begin by applying our estimates to the annual post-SPP birth rates and population counts to calculate the number of fewer births implied by our results. We then compare this to the actual number of female SPP recipients from 2015 to 2018. Our estimates imply that there were 23,878 fewer births to teenage mothers as a result of the introduction of SPP. This is larger than the 17,149 female SPP scholarship recipients during the same period. These calculations suggest that incapacitation or income effects from receiving the SPP scholarship cannot fully explain the effects we observe.

Next, returning to Figure 3, the timing of the effects we observe is also informative for the mechanisms driving the results. If incapacitation effects or income effects from receiving the scholarship primarily explain our results, we would expect to begin to see effects a few quarters after SPP's announcement when college enrollment began and scholarship funds were disbursed for the first cohort of SPP beneficiaries. However, we observe effects in the first quarter after SPP's announcement, which suggests that effects directly from receiving the scholarship are likely not the primary source of our estimated effects.

Finally, we use the more granular information in our data about the age of the father to learn more about what ages the fertility effects of SPP are concentrated. While our data only include mother's age in a range of years, the data do include father's age in integer years. Since the onset of SPP can have similar effects on the college opportunities for young men, we assess SPP's effect on teen fatherhood rates in Table 3, using the difference-in-differences specification from Equation 3 except with the age-specific fatherhood rates per 1,000 young men as the outcome.[16] The share of all fathers who are teens is lower than the share of mothers who are teens, and SPP's effects may not be the same between adolescent men and women. Nevertheless, the more disaggregated age-specific effects for adolescent men may still be informative for assessing which ages the fertility effects of SPP are concentrated and thus the class of mechanisms involved. For instance, larger effects for younger teenagers would suggest an important role for the indirect effects we previously described.

Table 3. Summary Difference-in-Differences Estimates of the Effect of *Ser Pilo Paga* on Teen Fatherhood

	IHS births per 1,000 men			
Age group:	All teens (15-19)	15-17	18	19
	(1)	(2)	(3)	(4)
$SPP \times Post$	-0.062***	-0.063**	-0.090***	-0.044*
	(0.022)	(0.032)	(0.027)	(0.024)
Observations	12,804	12,804	12,804	12,804
Treatment municipalities	541	541	541	541
Comparison municipalities	526	526	526	526
Pre-trends testing p-value	0.492	0.181	0.335	0.504
Pre-SPP share of teen fathers	100	25.7	33.6	40.7

Notes: This table presents summary difference-in-differences estimates using Equation 3 with the inverse hyberbolic sine (IHS) transformation of the number of births per 1,000 men in each age group as the outcome, because some municipalities in our main sample have zero births for certain male age groups. Column 1 presents the estimate for all male teens (15-19 years old). Column 2 shows the results for male teens 15-17 years of age. Columns 3 and 4 present the estimates for men 18 and 19 years of age, respectively. All estimates are weighted by the annual population of men in each municipality and age group. Standard errors are clustered at the municipality level and presented in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

[16] Since there are some municipalities that have zero births to teen fathers in some cells (particularly at young ages), we use the inverse hyperbolic sine transformation of the fatherhood rate as the outcome instead of the log transformation. In Table A2, we show our results are qualitatively similar using a Poisson model.

For all teenage men aged 15-19, we estimate that teenage fatherhood rates decreased by 6.2 percent in treatment municipalities relative to comparison municipalities after SPP was introduced, suggesting SPP had similar effects on parenthood between teenage men and women. Estimating effects separately by age, we find the smallest effects among 19 years olds, even though this group accounts for the largest share of births to teen fathers. Importantly, we find larger effects on fatherhood rates for teenagers aged 15-17, who are likely to have not yet received an SPP scholarship. Meanwhile, we find the largest effects among 18 year olds. 18 years olds can be high school graduates (and thus potential SPP scholarship recipients), but nearly 17 percent of SABER 11 test takers are 18 years old so many likely have not yet graduated high school.

Together, the results and observations above suggest that our main estimates largely consist of indirect effects, where new college-going opportunities created by SPP influenced teen fertility decisions for students before they were even able to benefit directly from the program. Our data limit us from precisely identifying the most relevant source of the possible indirect effects we describe. But, our results are consistent with Laajaj et al. (2022a) who find that SPP caused motivational effects on low-income students resulting in increased 9th grade test scores, years before eligibility for SPP is determined, throughout most of the test score distribution. These effects on 9th grade test scores materialized within the first year after SPP was announced, which supports the plausibility of the immediacy of our observed effects.

1.5.4 Heterogeneity Analysis

In this section, we explore how the fertility effects of SPP vary by pre-policy municipality characteristics. Specifically, in line with the idea that SPP represented a shock to post-secondary educational opportunities and opportunities for social mobility more broadly, we examine whether the program had different impacts based on the municipality's level of income inequality and students' college-going expectations before the policy's implementation.

Kearney and Levine (2014) show that rates of teen childbearing are closely related to income inequality, theorizing that feelings of greater economic hopelessness arise in the context of high income inequality, leading economically disadvantaged young women to perceive the opportunity costs of early childbearing to be low. Following this logic, we expect the fertility effects of SPP to be larger in magnitude in municipalities that ex-ante had higher levels of inequality, where the chances of economic mobility were arguably more limited. We use a Gini coefficient as a proxy of income inequality, measured in 2005 at the municipality level. We then separately estimate Equation 4 for those below the median (lower inequality) and above the median (higher inequality). These analyses also inform our main results where we largely interpret the fertility effects of SPP as an increase in economic opportunities.

We present the results of this analysis in Panel A of Table 4. Using subsets of municipalities necessarily changes the composition of treatment and comparison units used in the estimation. Thus, ensuring that the parallel trends assumption is still reasonably satisfied is important. Using the pre-trends test described in subsection 1.5, we cannot reject the null hypothesis that all pre-period coefficients for both subsets of municipalities are zero. Column 1 reproduces our core results for the overall sample of municipalities for which we have a measure of inequality (1,009 out of 1,067 in the main estimation sample), finding a similar estimate to the one in our main analysis. The estimated reduction in teen fertility in municipalities with higher levels of income inequality is 2.5 percentage points greater than in those with lower inequality (column 3 versus column 2). In fact, although negative, the estimated change in fertility in lower inequality municipalities is not statistically significant. However, we cannot reject the null hypothesis that these two coefficients are the same using a two-sided F-test (p-value = 0.295).

To assess the heterogeneous impacts of SPP by expectations of attending college, we utilize survey responses on the higher education expectations of a 10 percent random sample of SABER 11 test takers in 2013 and 2014 (pre-SPP). We use low-income (SISBEN 1 and 2) female students' responses to a question that asks about how likely they are to enroll in a higher education program immediately after finishing high school. At the municipality level, we calculate the share of respondents who indicate they are likely or highly likely to attend college. To guarantee that we have a reasonable number of students in each municipality, we only include municipalities for which we observe at least 5 percent of the test takers and at least ten students in this analysis. This limits the estimation sample to 658 municipalities.[17] Finally, we group municipalities by whether they are above or below the median share of students who meet this criteria and estimate Equation 4 separately for these two subsets of municipalities.

The triple difference coefficients are presented in Panel B of Table 4. First, we cannot reject the null that all pre-period coefficients for both subsets of municipalities are zero. Second, the overall coefficient in column 1 is similar to the one in our main analysis. We then estimate a reduction in fertility almost 3 percentage points larger (in magnitude) in municipalities where, before SPP, the expectations of immediate enrollment in higher education after high school graduation were smaller (column 2 vs column 3). This time we also cannot reject the null hypothesis that the two coefficients are the same using a two-sided F-test (p-value = 0.293).

We explore an additional source of heterogeneity. In Panel C of Table 4, we present results by splitting up municipalities according to the pre-SPP share of teen births to teen fathers. We find that the relative reduction in teen fertility is driven by municipalities where female teenagers tend to have children with other teenagers. In conjunction with the results presented in Table 3, this suggests a reinforcement of incentives to avoid parenthood when potential fathers also face a positive shock to college-going opportunities, or alternatively, an increase in bargaining power for female teenagers in relationships with peers of similar age.

Given the previous results, we complement our heterogeneity analysis by presenting in Figure A9 a slightly more detailed version of the results in Table 4. There, we divide municipalities in deciles according to the level of each characteristic discussed before. After recasting the triple difference as a simple difference-in-differences by calculating the within-municipality difference between the (log) fertility rate of teens and non-teens, we use Equation 3 to estimate the triple difference reduction in fertility for each decile. Following Muralidharan and Prakash (2017), we then plot these estimates using a lowess regression smoothing of the triple difference coefficients on the average level of the characteristic for the

[17] Results are qualitatively similar when using the slightly more stringent requirement of keeping municipalities where we observe at least 10 percent of the students.

Table 4. Triple Difference Estimates of the Effect of *Ser Pilo Paga* on Teen Fertility Rates by Group of Municipalities

	Log fertility rate		
Panel A. Baseline income inequality			
Municipalities:	All	Below median	Above median
	(1)	(2)	(3)
$Teen \times SPP \times Post$	-0.054***	-0.029	-0.054***
	(0.011)	(0.019)	(0.015)
Observations	24,216	11,928	13,680
Treatment municipalities	511	293	248
Comparison municipalities	498	204	322
Pre-trends test p-value	0.966	0.486	0.833
Panel B. Baseline college-going expectations			
Municipalities:	All	Below median	Above median
	(1)	(2)	(3)
$Teen \times SPP \times Post$	-0.053***	-0.068***	-0.039***
	(0.013)	(0.023)	(0.014)
Observations	15,792	7,896	7,896
Treatment municipalities	389	179	210
Comparison municipalities	269	150	119
Pre-trends test p-value	0.949	0.276	0.350
Panel C. Baseline share of teen births from adolescent fathers			
Municipalities:	All	Below median	Above median
	(1)	(2)	(3)
$Teen \times SPP \times Post$	-0.057***	-0.002	-0.062***
	(0.011)	(0.016)	(0.016)
Observations	25,608	12,744	12,864
Treatment municipalities	541	234	307
Comparison municipalities	526	297	229
Pre-trends test p-value	0.985	0.532	0.915

Notes: This table presents summary triple difference estimates using Equation 4 for a subgroup of municipalities indicated in each column. In Panel A, income inequality is measured by the 2005 municipal Gini obtained from CEDE's municipal panel. The Gini is not available for some of the municipalities. In Panel B, we classify municipalities according to the pre-SPP measure of college-going expectations described in the text. This measure is not available for some of the municipalities. In Panel C, the share of teen births from adolescent fathers corresponds to the average of 2008-2013. Two-sided p-value from an F-test of equality, H_0 (2) = (3): Panel A p-value = 0.295; Panel B p-value = 0.293; Panel C p-value = 0.009. All estimates are weighted by the annual population of women in each municipality and age group. Standard errors are clustered at the municipality level and presented in parentheses. (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

deciles. By thoroughly exploiting the variation in the measures of income inequality and college-going expectations, this exercise helps us confirm the takeaways from Table 4. Together, this set of results support the idea that SPP represented a bigger shock to economic opportunity in municipalities with greater ex-ante income inequality and with greater perceived limited opportunities for tertiary education.

1.6 Robustness

This section overviews a series of analyses that assess the robustness of our preferred triple difference results to alternative definitions of treatment and comparison units, possibly confounding events, and other sensitivity checks.

1.6.1 Alternative Definitions of Treatment and Comparison Units

We showed in subsection 1.5 that our main results are not dependent upon a convenient splitting of the sample into treatment and comparison municipalities at the median of SPP eligibility. Table A3 also shows that these results hold when using a continuous version of our treatment intensity measure, excluding capital cities from the sample, using the inverse hyperbolic sine (IHS) transformation of fertility rates as the outcome, and using fertility rates in levels as the outcome (estimated with either OLS or Poisson models). Both the IHS transformation and using the rates in levels allow us to keep the few municipalities with zero birth counts in our estimation. We build on this here by exploring additional definitions of treatment and control units using alternative sources of variation. Combined, these analyses show that our main results are not solely dependent upon our use of initial municipality-level SPP eligibility rates, or using women aged 25-29 as a comparison group.

First, we estimate a triple difference model similar to Equation 2 which replaces SPP^*_m with an indicator for whether a municipality is below the median distance (closer) to the nearest SPP-eligible institution.[18] Thus, this empirical approach does not rely at all upon a municipality's

[18] See Figure A4 for a map of where SPP-eligible institutions are located in Colombia. We calculate distance to institutions using only the initial set of institutions that were SPP-eligible at the start of the program.

initial SPP eligibility rate. We plot the event study estimates from this analysis in Figure A10. The results show that the age group differentials in log birth rates trend similarly, with no clear pattern, between municipalities that are closer and farther from SPP-eligible institutions from 2008 through 2014. After the introduction of SPP, however, these trends begin to diverge, with birth rates declining in municipalities closer to SPP-eligible institutions relative to those farther away. The summary estimate, shown in column 2 of Table A4, implies that municipalities closer to SPP-eligible institutions experienced a 9.6 percent decrease in teen fertility rates after SPP was introduced relative to municipalities that are farther away.

Second, we estimate a triple difference model that does not rely on using women aged 25-29 as the within-municipality control group. Here, we instead use young women whose birth record indicates their highest level of education completed being seventh grade or less as a comparison group. We replace $Teen_a$ in Equation 2 with an indicator for having completed eighth grade or more using only data on births to women aged 15-19.[19] The intuition is that any woman aged 15-19 whose highest grade completed is seventh grade or less is likely to have already dropped out of school. As school dropouts, these women are likely to be less affected by the introduction of SPP than women aged 15-19 who have completed eighth grade or higher.

The results of this approach are reported in Figure A11. Panel (a) shows, separately for treatment and comparison municipalities, the trends in the differential in log number of births between women age 15-19 who have completed eighth grade or higher and those who have completed only seventh grade or lower.[20] The plot shows that births to women with eighth grade or higher education is increasing over time relative to women with less than eighth grade education. But, these differentials trend very similarly between treatment and comparison municipalities from 2008 to 2014. Starting in 2015, however, the trends begin to diverge between

[19] We also restructure the analysis dataset to have one observation per year per municipality per grade-level (i.e., eighth grade or above, or seventh grade or less).

[20] We use the number of births instead of a birth rate because we cannot reliably calculate the number of young women in each municipality with above or below an eighth grade level of education. Also, we use an inverse hyperbolic sine transformation instead of a log transformation since some municipalities have zero recorded births in these year by grade-level cells.

these groups of municipalities, where the trends in grade level differentials flatten in treatment municipalities but continue to increase in comparison municipalities.

Panel (b) plots the triple difference estimates that compares the changes in the number of births to women aged 15-19 over time between those completing eighth grade or more versus seventh grade or less, and between treatment and comparison municipalities. The results mirror the trends in Panel (a): grade level differences in births between treatment and comparison municipalities trend similarly through 2014, but begin to diverge significantly after SPP is introduced in 2015. The summary estimate, presented in column 3 of Table A4, implies that births to adolescent women who have completed at least eighth grade decreased by 10.7 percent in treatment municipalities relative to comparison municipalities. This analysis highlights that our results are not dependent upon using older women aged 25-29 as a comparison group.

1.6.2 Possible Confounding Events

Since our setting involves a single treatment time period, we are potentially vulnerable to events that happened simultaneously (or around the same time) as the introduction of SPP. Although, we note that to truly be a threat to identification, these simultaneous events would have to differentially affect women of the different age groups and be correlated with SPP eligibility rates. Nevertheless, we assess whether three events that occurred at a similar time might be driving our results: 1) the unilateral permanent ceasefire by the Revolutionary Armed Forces of Colombia (FARC, from the Spanish acronym) in December 2014 as part of the by then ongoing peace process between the guerrilla group and the Colombian government, 2) the Zika virus epidemic, which occurred from October 2015 to July 2016, and 3) the *Jornada Única* initiative, which gradually transitioned some public secondary schools that were operating half-day shifts into full school days beginning in 2015.

For each of these possibly confounding events, we re-estimate our main specification using only a subset of municipalities that were likely unaffected by the relevant event. If these events are not driving our results, we would expect to see estimates based on these subsets of municipalities that are similar to our main estimates. We provide a full description of these analyses in Appendix C. and report these estimates in Table C1. Indeed, we consistently estimate large and statistically significant effects of SPP in each of these subsample analyses. We conclude that these three events cannot explain the effects we observe.

1.6.3 Other Sensitivity Checks

Although our empirical strategy does not rely on a staggered rollout design, the recent developments in the difference-in-differences literature have documented issues with estimating difference-in-differences designs with linear regressions and fixed effects specifications even with non-staggered binary treatments, which is the case of our specifications in subsubsection 1.4.2 (Borusyak et al., 2021; de Chaisemartin and D'Haultfœuille, 2022a,b). We implement the imputation estimator developed by Borusyak et al. (2021) and the difference-in-differences (DID) estimator proposed by de Chaisemartin and D'Haultfœuille (2022a) that are robust to treatment effect heterogeneity to estimate the models in Equation 2 and Equation 4. Results using these alternative estimators are reported in Figure A12 and Table A5. The triple difference estimate using the imputation estimator is 8.3 percent and for the DID estimator is 7.4 percent.

To assess whether the decline in teen fertility we observe in treatment municipalities is the result of pure chance, we perform a permutation test that randomly assigns municipalities to be treatment or comparison municipalities. We then compare our main estimate to a distribution of estimates across 5,000 randomly assigned groups of treatment municipalities. To do this, we use the randomization inference routine developed by Heß (2017) and the specification in Equation 4. We report the results in Figure A14. Reassuringly, we see that our main estimate is in the far left tail of the distribution of estimated triple difference coefficients. Also, to be sure our results are not driven by a small number of municipalities, we re-estimate our main specification while each time excluding municipalities in a single department (reported in Figure A13). We also estimate our main specification by excluding all municipalities that include a department's capital city (reported in column 2 of Table A3). The results from each of these regressions produce estimates that are very similar to our main estimates.

Finally, we perform a placebo-in-time strategy to further support the validity of the parallel trends assumption required for our estimates to have a causal interpretation. In Table A6, we use 2008-2014 data and estimate the same specification in Equation 4 pretending that SPP was introduced in years 2008-2013. The overall estimated effects after each of the placebo treatment years are always statistically insignificant and close to zero. The parallel trends assumption is an assumption about counterfactuals and, therefore, untestable. We have shown robust evidence that the differentials in fertility rates between younger and older women were not trending differently among treatment and comparison municipalities before SPP, which adds support that this would have been the case during the post-period had the program not been introduced.

1.7 Conclusion and Discussion

In this paper, we study the teen fertility impacts of *Ser Pilo Paga*, Colombia's generous college financial aid program for high-achieving, low-income students. After the 2014 introduction of the program, we find that teen fertility rates decreased by about 6 percent in municipalities more affected by SPP relative to less affected municipalities. Due to SPP being a nationally implemented policy, these estimates are necessarily relative effects. The *total* effects of SPP on teen fertility rates nationwide are likely to be larger than the relative estimates indicate.

While our data limit us from precisely identifying the mechanisms driving our results, our analyses point to effects largely coming from behavioral responses prior to students going to college, potentially including channels such as motivational effects, increased opportunity costs, and/or peer effects. We also find larger effects of SPP in areas where the pre-SPP levels of income inequality were greater. This is consistent with Kearney and Levine (2014) who present empirical and theoretical evidence that suggests inequality—and the "economic hopelessness" that inequality cultivates—explains a large share of the variation in teen childbearing rates. Placing our estimates in the context of existing research on the determinants of teen fertility in similar settings, we find that SPP's effects on teen fertility is smaller than *Familias en Acción*—Colombia's conditional cash transfer program—(Attanasio et al., 2021), but larger than a Chilean reform that lengthened the school day from half to full-day shifts (Berthelon and Kruger, 2011).

Our results suggest that increasing future economic opportunities for young women can lead to meaningful reductions in teen fertility, consistent with some of the policy considerations discussed by Kearney and Levine (2012) in the context of the United States. Prior to SPP, Colombia was characterized by large socio-economic gaps in college enrollment due to severe financial constraints, low access to credit, and high college tuition costs. We posit that, in countries with high inequality, college financial aid programs like SPP that decrease inequality of opportunity can have behavioral effects on teen childbearing and perhaps other outcomes. The characteristics of SPP—namely its generosity, salience, and simplicity—would seem to be important in accounting for the far-reaching impacts of the program, which is consistent with the college financial aid literature more broadly (e.g., Bettinger, Long, Oreopoulos and Sanbonmatsu (2012); Dynarski, Libassi, Michelmore and Owen (2021)).

In 2018, under a new presidential administration, the Colombian government announced that the *Ser Pilo Paga* program would no longer accept new beneficiaries and the program would be replaced. The program gained controversy during its four years due to its high cost to the government and the fact that most SPP beneficiaries attended private institutions. Our findings illustrate important indirect benefits of SPP. While such benefits alone may not justify the program's costs, they should be included in a full accounting of the program's costs and benefits, including the effects on college enrollment outcomes documented in Londoño-Vélez et al. (2020a).

Chapter 2: College Financial Aid and the Gender Achievement Gap: Evidence from Colombia

2.1 Introduction

Although female students in many high-income countries tend to perform better than their male peers in the reading component of different standardized tests, there is a well-documented average gender gap in test scores in favor of men in crucial subjects, particularly in mathematics and science, that typically worsens at the top of the achievement distribution and over the schooling cycle (Fryer and Levitt, 2010; Bharadwaj, De Giorgi, Hansen and Neilson, 2016; OECD, 2019). For example, Fryer and Levitt (2010) and Husain and Millimet (2009) present evidence of these gender gap patterns in the context of the United States, while Le and Nguyen (2018) do the same in the Australian context.[21] The overall panorama for women's relative performance in middle-income economies tends to be similar. In the context of Latin America and the Caribbean (LAC), a region with relatively marked gender gaps favoring men, Bharadwaj et al. (2016) for Chile and Abadía and Bernal (2017) and Muñoz (2018) for Colombia report evidence of these patterns.

Figure D1 presents a general overview of the country-level average gender gap in mathematics for a sample of approximately 100 high- and middle-income countries around 2014 using information from the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS). The data come from the Harmonized Learning Outcomes (HLO) database introduced by Angrist, Djankov, Goldberg and Patrinos (2021). The gender gap is measured as the ratio of the average male score to the average female score in standardized tests taken by students in secondary education. A value greater than one indicates that men outperform women on average. The plot shows that the male-favoring

[21] For the US, more recent results coming from the 2019 National Assessment of Educational Progress (NAEP) mathematics assessment show that male students in grade 12 outperform their female peers by 0.09 standard deviations on average and by 0.18 standard deviations at the 90th percentile.

gender gap in mathematics is a widespread phenomenon, with roughly 67% of countries in the sample exhibiting ratios over the unit.

Since these gaps have potentially significant downstream effects on college enrollment, major choices, and wage trajectories for women, investigating how to reduce them is a policy-relevant issue. As noted repeatedly by policymakers, for example, these gaps have negative consequences for women's participation in science, technology, engineering, and mathematics (STEM) fields (OECD, 2019). This paper contributes to our understanding of the evolution of the gender gaps in performance by studying the role that the availability of college financial aid can play in an environment with limited existing credit opportunities and large income-based gaps in college enrollment. The main result of the paper is that increasing financial aid for college can help reduce the gender gap in standardized tests among high school students.

The literature exploring the causes of the gender achievement gap is abundant. As exemplified by the works of Fryer and Levitt (2010) and Bharadwaj et al. (2016), both in high- and middle-income countries, a substantial portion of the achievement gap between men and women cannot be explained by differences in the observable characteristics and behaviors of male and female students, their parents or their teachers. The significant variation of the gap across and within countries, over time within countries and over the schooling cycle within cohorts of students, however, favors explanations related to broader societal norms and economic factors that tend to put women at a disadvantage relative to men over explanations related to differences in innate abilities of students of a given gender (Guiso, Monte, Sapienza and Zingales, 2008; Pope and Sydnor, 2010; Breda, Jouini and Napp, 2018; OECD, 2019; Gevrek, Gevrek and Neumeier, 2020).

In many low- and middle-income countries where men are more likely to drop out of the schooling system, particularly during secondary school, part of the average gap observed in mathematics has been linked to this phenomenon (Muñoz, 2018). Suppose men with lower performance in school are more likely to abandon the system than women of similar characteristics. In that case, the male-favoring average gap in standardized tests administered in lower- or upper-secondary school might partially reflect these compositional differences in the sample of test takers. Muñoz (2018) shows both descriptive evidence of this phenomenon across countries and causal evidence using policy-induced variation in the female-to-male ratio of test takers in the end-of-high school exam caused by the introduction of a conditional cash transfer program in Colombia that disproportionally incentivized female students to stay in high school during the early 2000s.

In this paper, I study how a dramatic expansion of financial aid for college that took place in Colombia in 2014 affected the gender achievement gap in standardized tests. The *Ser Pilo Paga* program (roughly translated as "Being a Good Student Pays Off") launched by the Colombian government to increase the college-going opportunities of top-performing low-income students benefited four cohorts of high school students from 2014 to 2017. Before the implementation of the policy, male students outperformed women on average by approximately 0.13 standard deviations in the end-of-high school standardized test. However, the gap widened to as much as 0.30 standard deviations at the 95th percentile of the test scores distribution. Consequently, women were underrepresented at the top of the *overall* distribution of test scores and at higher-quality, more selective universities. Four years after the announcement of *Ser Pilo Paga*, the gap at the top had decreased to 0.23 standard deviations. Using a group of students proxied to be from a non-low-income background as a comparison group and an extension of Athey and Imbens (2006)'s non-linear difference-in-differences model, I show evidence that the expansion of financial aid for high-achieving students caused a reduction in the gender gap at the top by between 9 and 13 percent, beginning around the 75th percentile of the distribution of test scores, partially explaining the observed change in the gap among top performers. I furthermore show that these estimates are not due to differential changes in observable characteristics of male and female test takers over time.

Even though the *Ser Pilo Paga* program did not have any explicit gender component, I present evidence that these results are due to a stronger response of female students when faced with the newly available college-going opportunities relative to men. I offer a theoretical framework to help rationalize these results and show that if exerting additional effort becomes too costly for students with initially higher scores, women will respond more strongly to the increased financial aid.

This paper contributes to the literature in several ways. First, it adds to the literature studying the causes of the gender achievement gap in standardized tests among secondary school students by providing evidence of the role that the availability of college financial aid has on the gap, given the human capital decisions students make in high school and the labor market conditions they face (or expect to face) after high school. As initially documented by Bernal and Penney (2019) and more comprehensively by Laajaj, Moya and Sánchez (2022a), the unprecedented opportunities created by the introduction of *Ser Pilo Paga* (SPP) caused an increase in effort and learning among low-income students materialized in higher scores in the end-of-high school exam that partially qualified students to receive the scholarship provided by the program. Laajaj et al. (2022a)'s argument is in part that the pre-SPP environment characterized by limited college-going opportunities for low-income high school students was inducing them to leave *potential* human capital accumulation on the table. The results presented here additionally imply that the gender achievement gap should shrink as barriers to access to higher education decrease. This contribution is related to the work of Muñoz (2018), in that it relates the gender achievement gap to broader economic factors affecting the labor market returns of educational investments during high school. I, however, focus on a different, albeit related, margin: opportunities for college attendance instead of dropping-out decisions during high school (given differential outside options). Second, this paper adds to the literature documenting the gendered effects of educational interventions (e.g, Avitabile and de Hoyos (2018) and, for a review, Evans and Yuan (2021)). As emphasized by Evans and Yuan (2021), I provide evidence that *general* educational interventions (not targeting female students specifically) can effectively improve the relative position of women. Finally, this paper adds to the growing literature related to the effects of the *Ser Pilo Paga* program. SPP had significant—immediate and persistent—positive effects on college enrollment among low-income students (Londoño-Vélez, Rodríguez and Sánchez, 2020a; Londoño-Vélez, Rodríguez, Sánchez and Álvarez, 2023), positive (*ex ante*) motivational effects on test scores as mentioned before (Bernal and Penney, 2019; Laajaj et al., 2022a), and indirect effects contributing to a reduction in teen fertility rates (Bloem and Villero, 2023) and improved awareness of income inequality and preferences for redistribution among high-income students (Londoño-Vélez, 2022). I extend this body of work by studying the distributional impacts of SPP on the gender achievement gap. Unlike Laajaj et al. (2022a), this paper does not focus on the *socioeconomic* achievement gap between low-income and non-low-income students. The average achievement gap in test scores between men and women among Colombian students from the cohorts of 2013 and 2014 (pre-SPP) was around 45% of the socioeconomic achievement gap, so it is a meaningful and important dimension of inequality that grants further investigation.

The rest of this paper is organized as follows: In the next section, I provide an overview of the *Ser Pilo Paga* program and describe the patterns of the gender achievement gap in Colombia for the last cohort of students prior to the program's announcement. Next, I introduce a theoretical framework to help understand the differential response of women to financial aid in the third section. Following this, the fourth section will explain the data sources and how the groups of *low-income* and *non-low-income* students were defined for the analysis. Then, in section five, I will discuss the identification and estimation approaches used. I present the key empirical findings in the sixth section. In section seven, I present some robustness tests of these results. Finally, the last section concludes.

2.2 Background

2.2.1 The Ser Pilo Paga Program

The Colombian government announced the *Ser Pilo Paga* (SPP) program in October of 2014, two months after the 2014 cohort of high school students, the first to be eligible for the program, had taken the end-of-high school standardized exam. The program's objective was to offer financial support to high-performing, low-income students to attend any post-secondary program in Colombia at a Higher Education Institution (HEI) with a High Quality Accreditation, a distinction granted by the Ministry of Education upon the request of the HEIs and after successfully passing a pre-defined evaluation process. The program provided a full scholarship for tuition and a stipend to help cover other attendance-related costs. Regarding eligibility, the program had both a merit and a needs component.

Students had to score above a year-specific threshold on the standardized high school exit exam, SABER 11. SABER 11 scores have a considerable impact on college admissions, as approximately 70% of HEIs take them into account when evaluating applicants for admission (OECD, 2016). The SPP threshold was set around the 91st percentile of the overall distribution for the first two cohorts benefiting from SPP, and the cut-off for the last two cohorts increased to around the 95th percentile. However, the raw score required to be eligible for SPP increased year after year from 310 to 348 points (out of 500) from 2014 to 2017, given the already documented overall effects of the program on test score performance at the top of the distribution and the annual target number of scholarships the government had (ten thousand, on average).

Eligibility on the needs component was determined by the score of the student's household on the SISBEN, a household survey administered to potential beneficiaries of social programs in the country. The SISBEN score had to be below a certain threshold that varied with the student's residence area (up to main urban areas, rest of urban areas, and rural areas). At the moment of its announcement, around 52% of the SABER 11 test takers qualified for the scholarship on needs grounds.

SPP benefited almost 40,000 students over four cohorts of high school students. The 2017 cohort was the last one benefiting from the scholarship, since in September of 2018, under a new president, the Colombian government announced that the program would not be taking any new beneficiaries due to the high burden of the program on the national higher education budget.[22]

[22] The 2018 cohort was the last to take the SABER 11 exam with the expectation of obtaining an SPP scholarship since they took the test in August of 2018 before the national government formally announced the end of the program.

The income-driven gaps in access to higher education, and in particular high-quality higher education, among top-performing students were high before SPP. As documented by Londoño-Vélez et al. (2020a), the program successfully eliminated these gaps.

2.2.2 The Gender Achievement Gap in Colombia

In 2014, the observed gender gap in average performance on the SABER 11 test was 0.13 standard deviations in favor of male students. As in many other countries, there was substantial within-country variation in the gap across departments (states) and along the distribution of test scores. These patterns have been carefully documented by Abadía and Bernal (2017).[23] In this section, I present a summary of these distributional patterns and discuss how they relate to women's access to higher education. Following the literature, I measure the gender achievement gap at each quantile of the distribution of test scores as the horizontal difference between the marginal distributions defined by the gender of the student (Husain and Millimet, 2009). More precisely, I estimate the following equation using a traditional quantile regression for quantiles $q \in \{5, 10, \ldots, 95\}$:

$$\text{SABER 11 score}_i = \alpha_{0,q} + \alpha_{1,q} Male_i + \varepsilon_i. \tag{5}$$

In Equation 5, $Male_i$ is an indicator for male students and $\alpha_{1,q}$ represents the gender achievement gap at quantile q. For example, a gap of 0.21 at the 75th percentile, $\widehat{\alpha}_{1,75} = 0.21$, means that the 75th percentile of the distribution of test scores for men is greater by 0.21 standard deviations than the 75th percentile of the distribution of test scores for women. Following Husain and Millimet (2009), I loosely refer to $\alpha_{1,q}$ as the quantile treatment effect (QTE) of gender. I present estimates for unconditional gaps (without adjusting for covariates) and adjusted gaps, in both cases using Equation 5.[24] To adjust for covariates, I follow Husain and Millimet (2009) and

[23] Figure D1 offers a cross-country comparison of the average gender gap in secondary education. The Latin American average and Colombia's position are included for reference.

[24] One can adjust for covariates in different ways. The most restrictive way is to directly include them in Equation 5. This approach creates quantiles of residualized outcomes, which affects the interpretation of the gaps in terms of the

reweight each observation by the inverse of the propensity score as follows

$$\omega_i = \frac{Male_i}{\Pr(Male = 1|X)} + \frac{1 - Male_i}{1 - \Pr(Male = 1|X)}, \tag{6}$$

where $\Pr(Male = 1|X)$ is estimated parametrically using a probit model.[25] The vector X includes indicators for the student's age, mother's education, father's education, family size, attending a public school, school schedule (full-day, afternoon, evening, weekends), and attending a rural school.

Figure 4 shows the raw and adjusted gender gaps for 2014. The observed gaps at the top of the distribution are striking. At the 90th percentile, the observed gap is 0.27 standard deviations, two times the average gap. The adjusted gaps tend to be bigger for most of the distribution except at the top, where they are virtually the same as the unadjusted ones. As shown by Abadía and Bernal (2017) using data for the same period, individual, family, and school characteristics do not dramatically change the general pattern of the gaps. As a consequence of these large gaps, women were underrepresented at the top of the *overall* distribution of test scores. While women represented 55% of the test takers, they only accounted for 45% of the students scoring at the top decile.

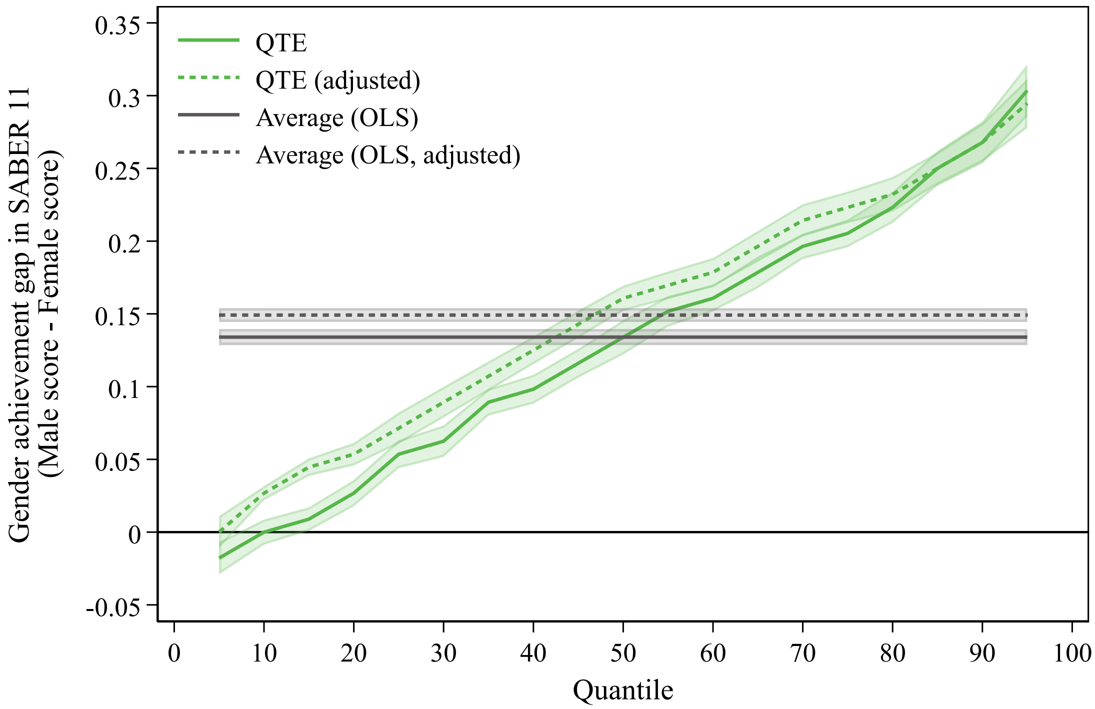
I estimate the following equation to shed light on the consequences of the previously described gaps for college enrollment:

$$\text{HEI Enroll}_i = \beta_0 + \beta_1 Male_i \,(+\, \beta_2 SABER11_i + \beta_3\,(Male_i \times SABER11_i)) + X_i\Gamma + \varepsilon_i. \tag{7}$$

unconditional distributions. An alternative would be to reweight the distributions by the inverse of a propensity score capturing the probability of observing a student of a particular gender given their characteristics and then estimate the gaps using these reweighted distributions as in the unadjusted case. This is the approach followed by Husain and Millimet (2009). More advanced methods include the estimation of quantile treatment effect models using Recentered Influence Functions (RIF) in a first step (with or without reweighting) and then adjusting for covariates in a second step through an OLS regression with the RIF values as the outcomes (Firpo, Fortin and Lemieux, 2009; Firpo and Pinto, 2016; Rios-Avila, 2020).

[25] For inference, ω_i is re-estimated with every bootstrap replication.

Figure 4. Gender Achievement Gap in Colombia (2014)



Notes: This plot presents estimates of the gender achievement gap in SABER 11 test scores in 2014. Test scores are standardized using the overall sample mean and standard deviation. For each quantile $q \in \{5, 10, \ldots, 95\}$, the green lines report $\widehat{\alpha}_{1,q}$ from a quantile regression using Equation 5. The dashed green line reweights each observation using estimates of the weights in Equation 6 as explained in the text. The shaded green areas represent pointwise 95% confidence intervals based on 5,000 bootstrap replications stratifying by gender. The gray lines present the average gap from OLS regressions using Equation 5, with the 95% confidence intervals also based on 5,000 bootstrap replications stratified by gender for comparability.

Using data from Londoño-Vélez et al. (2020b) for the 2014 cohort of high school students (who benefited from SPP but took the SABER 11 test without knowledge of the program), column 1 in Table 5 shows that male students were 1.7 percentage points (or 9 percent) more likely to enroll in any college right after finishing high school than female students from similar backgrounds.[26] However, column 2 shows that this gap disappears (and even reverses) when you compare students with the same performance level on the SABER 11 exam. Columns 3 and 4 indicate that men were also more likely to enroll in high-quality universities (1.5 p.p. or 25%), but that, again, the advantage disappears after controlling for the performance on the standardized

[26] Table D1 in the appendix shows these estimates using the 2013 cohort of high school students. The main takeaways are the same.

high school exit test. Columns 5 and 6 re-estimate columns 3 and 4, restricting the sample to the subset of students who enrolled in any college. Overall, this exercise shows that conditional on achievement, post-secondary enrollment for women and men is very similar, but women are underrepresented at the top of the distribution of test scores and therefore at better and more selective colleges. These gaps also manifested in the number of female students who received a SPP scholarship over the four cohorts of the program (43% of the total). Among the first cohort of potential SPP beneficiaries, women represented 57% of the SISBEN eligible students (close to their overall share of the student population), but only 45% of the SABER 11 eligible (way below their share).

Table 5. Test Scores' Role in the Gender Gap in Higher Education Enrollment

	Any HEI		High Quality HEI		High Quality HEI (Any HEI = 1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Male (×100)	1.747***	-0.487***	1.499***	-0.008	5.041***	1.257
	(0.103)	(0.172)	(0.066)	(0.052)	(0.278)	(1.308)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
SABER 11		Yes		Yes		Yes
Observations	540,889	540,889	540,889	540,889	109,374	109,374
Female mean (%)	19.4	19.4	5.9	5.9	30.5	30.5
Gap (%)	9.0	-2.5	25.3	-0.1	16.5	4.1

Notes: This table shows estimates from Equation 7. It uses data for the 2014 cohort of SABER 11 test takers from Londoño-Vélez et al. (2020b). The outcome is an indicator variable for immediate post-secondary enrollment in a higher education institution (HEI). High-quality HEIs are those with a government-granted High Quality Accreditation. Controls include indicators for a proxy of socioeconomic status, age, mother's education, father's education, family size, attending a public school, and the school schedule. The SABER 11 control is included flexibly with indicators for performance deciles. Robust standard errors are shown in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

2.3 Theoretical Framework

In this section, I introduce a simple model that relates the level of effort that students of different genders put into the end-of-high school exam that partially determines their probability of entering college to the availability of college financial aid. The main takeaway of the model is that since men performed better than women throughout most of the distribution of test scores before SPP and particularly at the top (the target of the program), under reasonable assumptions, the expected net gain of exerting additional effort in response to increased financial aid will be smaller among men. This translates into a decrease in the gender achievement gap after financial assistance is expanded.

The model presented below borrows elements from the models in Becker, Hubbard and Murphy (2010), Roland and Yang (2017), Muñoz (2018), and Laajaj et al. (2022a). In particular, the general problem of a student trying to decide the optimal level of effort to put into a test used for college admissions is similar to the one in Roland and Yang (2017), even though their context and motivation are substantially different. I deviate from their basic specification of the expected returns to college to allow for gender differences in net returns and to introduce the financial aid component that is critical to the problem in the current setup. In this regard, I broadly follow Becker, Hubbard and Murphy (2010) and Muñoz (2018) in modeling the decisions of the students as a function of the gender-specific labor market options they face. In contrast with Muñoz (2018), however, the model here does not focus on the marginal student deciding to stay in high school or not but instead focuses on students for which finishing high school was optimal and who have to decide how much effort to put into the end-of-high school exam that will partially determine their college prospects. Finally, the model incorporates the existence of a test score production function and a test score threshold like the model in Laajaj et al. (2022a). In the present model, this threshold is related to the overall chances of getting into college and not to the chances of receiving a scholarship. Also different from Laajaj et al. (2022a), the test score production function here is assumed to be only a (non-linear) function of effort and to depend on the student's ability indirectly through the optimal choice of effort.

2.3.1 Setup

There are female and male students, indexed by $g \in \{f, m\}$. For both genders, the student's score on the end-of-high school exam (S) is a stochastic function of the student's effort e

and it is given by

$$S = S(e) + \tilde{e}, \tag{8}$$

where $\tilde{e} \sim F^{\tilde{e}}$ is the random component of the outcome S and it is assumed to have zero mean. $S(e)$ is differentiable and strictly increasing in effort, and so $S_e > 0$. Students choose their level of effort to maximize their expected net gain from attending college, considering their probability of admission and the costs of exerting effort. An exogenous threshold, $\bar{S}$, exists above which students get into college. The probability of getting into college is then given by

$$\pi(e) = \Pr\left(S(e) > \bar{S}\right) = 1 - F^{\tilde{e}}\left(\bar{S} - S(e)\right). \tag{9}$$

Like in Roland and Yang (2017), I assume that the probability of getting into college increases with effort but at a decreasing rate. That is, $\pi(e)$ is an increasing and strictly concave function of effort. I also assume $\pi(e)$ is differentiable as necessary. If, for example, the random component in S is uniformly distributed, $\tilde{e} \sim \mathrm{U}(\tilde{e}^{-}, \tilde{e}^{+})$, a sufficient condition for $\pi(e)$ to meet these requirements is that $S_e > 0$ (assumed earlier) and $S_{ee} < 0$. So it would be sufficient for $S(e)$ to be a strictly concave function of effort.

Let W_g denote the present value of lifetime earnings that a student of gender g is expected to make if they enroll and finish college. Similarly, let w_g denote the present value of lifetime earnings that a high school graduate of gender g is expected to make if they do not enroll in college. If students enroll in college, they must pay for tuition and other college-specific costs in the amount of T plus the cost of raising these funds captured by an interest rate r_g, which is allowed to vary by gender. The government can partially subsidize the cost of tuition. The policy parameter $\theta \in [0, 1]$ represents the portion of tuition the government covers. Finally, $C(e, a)$ represents the cost of exerting effort for a student with innate ability a. As in Roland and Yang (2017) and Laajaj et al. (2022a), I assume that $C(e, a)$ is differentiable, increasing and strictly convex in e, so $C_e > 0$ and $C_{ee} > 0$. Following Becker et al. (2010) and Roland and Yang (2017), I additionally assume that students with greater innate ability face a smaller marginal cost of effort, and so $C_{ea} < 0$.[27] I do not make functional forms assumptions about the distribution of ability in the population beyond assuming that it does not depend on gender.

Combining all the previous elements, the problem that a student of gender g solves is then given by

$$\max_{e} \; \pi(e) \left[W_g - (1 + r_g)(1 - \theta)T\right] + [1 - \pi(e)]\, w_g - C(e, a). \tag{10}$$

2.3.2 The Equilibrium Effort Level and the Gender Gap in Performance

For students choosing a positive level of effort (and dropping the g subscript), the first-order condition (FOC) for the problem in Equation 10 is

$$\pi_e[W - w - (1 + r)(1 - \theta)T] = C_e. \tag{11}$$

Equation 11 implies that students choose the level of effort that equalizes their expected marginal benefit of exerting effort (left-hand side) with the marginal cost (right-hand side). The expected marginal benefit is the change in the probability of getting into college when effort increases times the net benefit of attending college (i.e., the value of the increased chance of attending college).

Proposition 1. *Let $B \equiv W - w$ denote the lifetime college earnings premium. Under the assumptions laid out before, the optimal level of effort $e^* = e(B, r, \theta, T, a)$ is increasing in the college earnings premium and the level of college financial aid, and is decreasing in the difficulty of raising tuition funds (proxied by the interest rate). That is, $\frac{de^*}{dB} > 0$, $\frac{de^*}{d\theta} > 0$, and $\frac{de^*}{dr} < 0$.*

[27] The model in Becker et al. (2010) deals with this as a complementarity in the production function between ability ("cognitive skills") and schooling. Note, however, that their model is about gendered decisions about investments in college education.

The proof for Proposition 1 comes from totally differentiating Equation 11 with respect to B, θ, and r:

$$\pi_{ee}[B-(1+r)(1-\theta)T]de^* + \pi_e dB = C_{ee}de^*$$
$$\frac{de^*}{dB} = \frac{-\pi_e}{\pi_{ee}[B-(1+r)(1-\theta)T]-C_{ee}} > 0 \tag{12}$$

$$\pi_{ee}[B-(1+r)(1-\theta)T]de^* + \pi_e(1+r)Td\theta = C_{ee}de^*$$
$$\frac{de^*}{d\theta} = \frac{-\pi_e(1+r)T}{\pi_{ee}[B-(1+r)(1-\theta)T]-C_{ee}} > 0 \tag{13}$$

$$\pi_{ee}[B-(1+r)(1-\theta)T]de^* - \pi_e(1-\theta)Tdr = C_{ee}de^*$$
$$\frac{de^*}{dr} = \frac{\pi_e(1-\theta)T}{\pi_{ee}[B-(1+r)(1-\theta)T]-C_{ee}} < 0. \tag{14}$$

The signs of the previous expressions follow from the assumptions that $\pi_e > 0$, $\pi_{ee} < 0$, and $C_{ee} > 0$. □

Proposition 2. *If the lifetime college earnings premium of men is greater than the lifetime college earnings premium of women, $B_m > B_f$, the equilibrium performance of women in the end-of-high school exam will be lower than the performance of men, $S_m > S_f$.*

From Proposition 1 $\frac{de^*}{dB} > 0$, so $B_m > B_f$ implies $e_m^* > e_f^*$. Now since $S_e > 0$, then $S(e_m^*) > S(e_f^*)$. Finally, it follows that $S_m = S(e_m^*) + E(\tilde{e}) > S(e_f^*) + E(\tilde{e}) = S_f$, where the expectation is taken over $F^{\tilde{e}}$ and so $E(\tilde{e}) = 0$. □

2.3.3 The Change in the Equilibrium Effort Level in Response to Financial Aid

I have shown so far that male students will outperform female students in equilibrium under the assumption that men enjoy a greater college earnings premium. This assumption is likely consistent with the Colombian evidence. Women not only have lower labor force participation rates but also tend to receive lower wages at all parts of the distribution of wages, with a male-favoring gap that is higher at the top than at the median of the wage distribution (Badel and Peña, 2010; Emiliani and Barón, 2012; World Bank, 2019). Furthermore, women seem to have a lower internal rate of return to college education than men (García-Suaza, Guataquí, Guerra and Maldonado, 2014).[28]

Even though I am relying on the differences in gender-specific lifetime earnings premiums, the model discussed before also allows for part of the equilibrium performance differential to emerge because of gender differences in the cost of tuition and other college-specific factors driven by different costs of obtaining these funds (as proxied by the interest rate). Women may face discrimination in credit markets, or gender social norms could make families less willing to pay for their daughters' education.[29]

I now explore how the magnitude of the financial aid-induced change in the equilibrium level of effort depends on the college earnings premium. I show that as long as the expected gain from increasing effort is limited, as detailed below, female students will respond more positively to a tuition subsidy relative to male students.

Proposition 3. *Under the assumption that* $\pi_{eee}[B-(1+r)(1-\theta)T]-C_{eee}\leq 0$*, the change in the optimal level of effort in reaction to an increase in financial aid decreases with the college earnings premium. That is, if* $\pi_{eee}[B-(1+r)(1-\theta)T]-C_{eee}\leq 0$*, then* $\frac{d\xi^{\theta}}{dB}<0$*, with* $\xi^{\theta}\equiv\frac{de^*}{d\theta}$.

[28]Additionally, data for 2019 from the Colombian National Department of Statistics (DANE, from the Spanish acronym) show that the male-to-female ratio of the raw difference between the average monthly earnings of a college graduate and the average monthly earnings of a high school graduate was 1.3. See `https://www.dane.gov.co/files/investigaciones/notas-estadisticas/nov-2020-brecha-salarial-de-genero-colombia.pdf`.

[29]See de^*/dr in Proposition 1. There is evidence for the US that women are more likely to take on student loan debt (Miller, 2017). There is also some evidence for Peru that female students are disproportionately more sensitive to the costs of attending post-secondary education (Molina, Santa María and Yamada, Forthcoming). In Colombia, some indirect evidence suggests that women are also more likely to take out student loans. While they represented around 52% of the first-year undergraduate enrollment in 2015, they accounted for 57% of the new loan recipients among undergraduate students granted by the Colombian Institute for Educational Credit and Technical Studies Abroad (ICETEX, from the Spanish acronym), the public institution administering students loans.

Differentiating Equation 13 with respect to B you get

$$\frac{d\xi^{\theta}}{dB} = \frac{G - H}{\left\{\pi_{ee}[B - (1+r)(1-\theta)T] - C_{ee}\right\}^2}, \tag{15}$$

where

$$G = -\pi_{ee}(1+r)T\frac{de^*}{dB}\left\{\pi_{ee}[B - (1+r)(1-\theta)T] - C_{ee}\right\} \text{ and}$$
$$H = -\pi_{e}(1+r)T\left\{\pi_{eee}[B - (1+r)(1-\theta)T]\frac{de^*}{dB} + \pi_{ee} - C_{eee}\frac{de^*}{dB}\right\}.$$

By concavity of $\pi(e)$, convexity of $C(e,a)$ in e and $\frac{de^*}{dB} > 0$ from Proposition 1, I have that $G < 0$. Since $G < 0$, a *sufficient* condition for $\frac{d\xi^{\theta}}{dB} < 0$ is that $H > 0$. Given $\frac{de^*}{dB} > 0$ and $\pi_e > 0$, a *sufficient* condition for $H > 0$ is then $\pi_{eee}[B - (1+r)(1-\theta)T] - C_{eee} \leq 0$. □

Intuitively, the condition $\pi_{eee}[B - (1+r)(1-\theta)T] - C_{eee} \leq 0$ implies that the expected *net marginal* benefit function $\pi_e[B - (1+r)(1-\theta)T] - C_e$ should be (weakly) concave around the optimal level of effort. For instance, the previous sufficient condition holds true for a quadratic cost function and a sufficiently smooth probability function $\pi(e)$ that can be well-approximated by a second-order approximation around $e = e^*$. In this last case, both $\pi_{eee} = 0$ and $C_{eee} = 0$, and

$$\frac{d\xi^{\theta}}{dB} = \frac{-\pi_{ee}(1+r)T\frac{de^*}{dB}\left\{\pi_{ee}[B - (1+r)(1-\theta)T] - C_{ee}\right\} + \pi_e(1+r)T\pi_{ee}}{\left\{\pi_{ee}[B - (1+r)(1-\theta)T] - C_{ee}\right\}^2} < 0. \tag{16}$$

More generally, conceptually, what is required is for the objective function to become sufficiently unresponsive to additional effort as the amount of effort increases. This can be explained through either the probability or cost function channels. For example, it could be that studying more does not necessarily lead to better results if there is a point where further studying becomes unproductive or that the additional effort required to obtain better results is unreasonably costly.

Corollary 1. *If the lifetime college earnings premium of men is greater than the lifetime college earnings premium of women, $B_m > B_f$, and Proposition 3 holds, the response of women to an increase in financial aid in terms of effort will exceed the response of men. That is,* $\xi_f^\theta \equiv \frac{de_f^*}{d\theta} > \frac{de_m^*}{d\theta} \equiv \xi_m^\theta$.

This follows directly from Proposition 3. If $\frac{d\xi^\theta}{dB} < 0$, $B_m > B_f$ implies $\xi_f^\theta > \xi_m^\theta$. $\square$

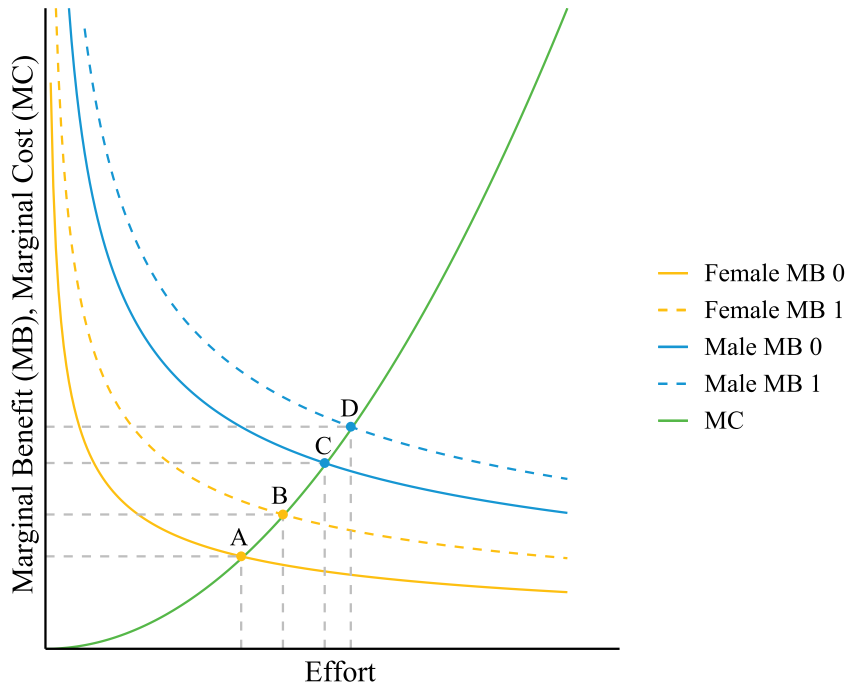
2.3.4 A Graphical Representation of the Differential Response to Financial Aid

Figure 5 shows a graphical representation of the main takeaways of the model up to this point. It presents the marginal benefit curves (MB) for male students (in blue) and female students (in yellow), under the assumption that men face a higher college earnings premium. The common marginal cost curve (MC) is depicted in green. Points A and C show an initial situation in which men are able to exert a higher level of effort in equilibrium, as discussed in Proposition 1 and Proposition 2. The figure then shows an upward shift of the marginal benefit curves representing an identical increase in the tuition subsidy of $\Delta\theta > 0$ for students of both genders that translates into an increase in the net return to college by $\Delta\theta(1 + r)T > 0$. Points B and D show the new equilibrium levels of effort, which are greater for students of both genders, since, as I showed before, $\xi^\theta \equiv \frac{de^*}{d\theta} > 0$. However, the increase in female effort (moving from A to B) is greater than the increase in male effort (moving from C to D). This follows from Proposition 3 and Corollary 1. At points A and C, it is more costly for men to exert additional effort in response to an increase in the net benefits of attending college because their marginal cost increases more rapidly (relative to women).

2.3.5 Differential College Earnings Premium, Ability and the Effort Response of the Marginal College-Goer

The analyses made before allow us to compare the response of male students with the response of female students *with the same level of ability*. However, given that the threshold $\bar{S}$ to get into college is the same for students of both genders and that the distribution of ability in the

Figure 5. Average Differential Effort Response to Increased Financial Aid



population is independent of gender, a higher proportion of men will cross the threshold $\bar{S}$ because men's equilibrium level of effort is higher than women's. That is, $F^{\tilde{e}}(\bar{S} - S(e_f^*)) \geq F^{\tilde{e}}(\bar{S} - S(e_m^*))$, given $e_f^* < e_m^*$ and $S_e > 0$. I now explore how the *marginal* student (i.e., the last student crossing $\bar{S}$) would respond to a change in financial aid.

Proposition 4. *Let $a_g^{\bar{S}}$ denote the ability of the marginal student of gender g crossing the college test score threshold $\bar{S}$. If the college earnings premium is higher for men than for women ($B_m > B_f$), the ability of the marginal women surpassing $\bar{S}$ is greater than that of the marginal men ($a_f^{\bar{S}} > a_m^{\bar{S}}$).*

Since the disturbance in S in Equation 8 has zero mean, $E(\tilde{e}) = 0$, there will be an expected level of effort $e^{\bar{S}} = S^{-1}(\bar{S})$ necessary to get into college. Given that the marginal student is also maximizing their expected net benefit of attending college, consider the FOC in

Equation 11 evaluated at $e^{\bar{S}}$ for students of both genders:

$$\pi_e|_{e=e^{\bar{S}}}[B_m - (1+r)(1-\theta)T] - C_e|_{(e,a)=(e^{\bar{S}},a_m^{\bar{S}})} = 0 \tag{17}$$

$$\pi_e|_{e=e^{\bar{S}}}[B_f - (1+r)(1-\theta)T] - C_e|_{(e,a)=(e^{\bar{S}},a_f^{\bar{S}})} = 0 \tag{18}$$

Subtracting Equation 18 from Equation 17, factoring $\pi_e|_{e=e^{\bar{S}}}$, and reorganizing:

$$\pi_e|_{e=e^{\bar{S}}}(B_m - B_f) = C_e|_{(e,a)=(e^{\bar{S}},a_m^{\bar{S}})} - C_e|_{(e,a)=(e^{\bar{S}},a_f^{\bar{S}})} \tag{19}$$

Because $\pi_e > 0$, if $B_m > B_f$, then it must be the case that $C_e|_{(e,a)=(e^{\bar{S}},a_m^{\bar{S}})} > C_e|_{(e,a)=(e^{\bar{S}},a_f^{\bar{S}})}$ for the previous equality to hold. But since C_e is a function of only e and a and $C_{ea} < 0$, at $e = e^{\bar{S}}$ it must be that $a_m^{\bar{S}} < a_f^{\bar{S}}$. $\square$

Figure 6. Differential College Earnings Premium and the Effort Response of the Marginal College-Goer

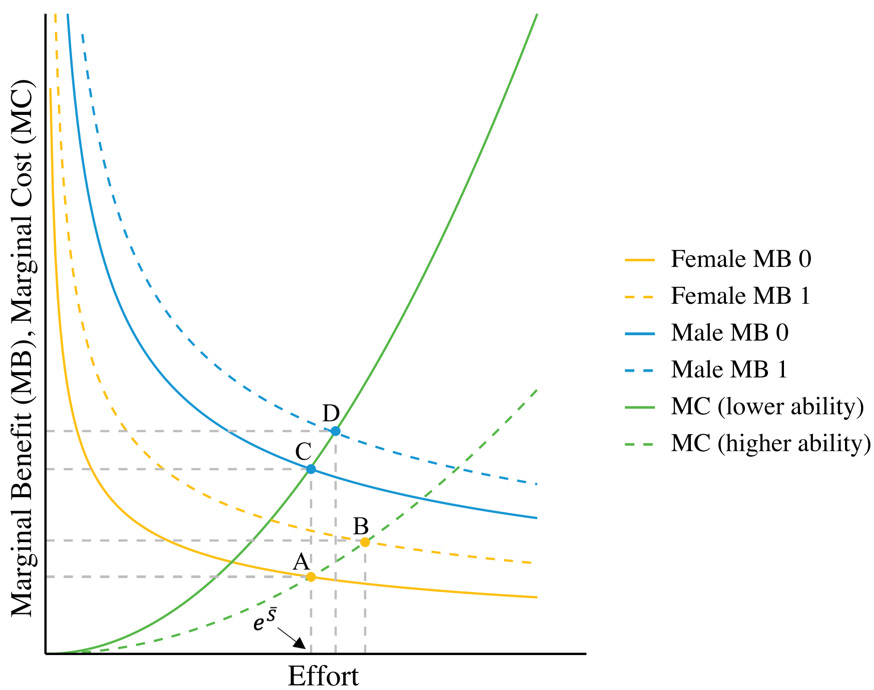


Figure 6 represents the situation of the marginal college-goer. Starting at points A and C, the marginal female student is exerting the same level of effort required to get into college as the marginal male (defined by $e^{\bar{S}}$). Given the differential in marginal benefits, this is only possible on this plane if the female marginal cost curve sits below the male marginal cost curve. The dotted green line then represents the marginal cost curve of a student with higher ability. Points B and D show the new equilibrium levels of effort after an identical increase in financial aid. The increase in women's effort level (moving from A to B) is greater than the increase for men (moving from C to D). Again, this is possible because the marginal woman is facing a smaller increase in the marginal cost (relative to the marginal man) due to the ability differential.

2.4 Data and Key Variables

2.4.1 Data Sources

I use administrative records from the Colombian Institute for the Assessment of Education (ICFES, from its Spanish acronym) containing information on the individual-level results of the universe of students taking the SABER 11 exam from 2011 to 2018.[30] The SABER 11 test is taken by nearly all high school seniors, independently of their plans of enrolling in post-secondary education. Universities and other higher education institutions use the test results as part of their admission criteria, either exclusively or with other elements like university-specific exams or interviews. As mentioned before, a significant share of colleges use the test as part of their admissions (OECD, 2016). The administrative records provided by ICFES also have information about individual and school characteristics supplied by the students and their schools when registering for the exam.

The college enrollment data for 2013-2014 used in subsubsection 2.2.2 come from Londoño-Vélez et al. (2020b). The information for the number of students eligible for SPP on the SISBEN margin by school discussed in subsubsection 2.4.2 comes from Laajaj et al. (2022b). Other additional sources of information used throughout the text are mentioned when first introduced.

[30] I use data for all students taking the test during the second semester of each year. The SABER 11 exam is administered by ICFES twice per year. Still, most students take the test during the second semester (typically in August/September), and those successfully enrolling in college start in January/February of the following year.

2.4.2 Proxy for Low-Income Status

The SPP program defined in a precise way the eligibility for the scholarship on the SISBEN margin: Students with a household-level SISBEN score above the pre-defined threshold were not eligible for the program. Students who did not have a SISBEN score were, in general, also ineligible. Ideally, I would use students non-eligible for SPP on the SISBEN margin as the comparison group. However, I do not observe the SISBEN score of the students, and I only have access to a self-reported proxy for SISBEN eligibility for two years after the program's announcement. To ensure that I use a reliable measure of socioeconomic status that can be used throughout the whole period of analysis, I leverage school-level data on the share of students eligible for SPP on the SISBEN margin in 2014 from Laajaj et al. (2022b). Figure D2 shows the distribution of this measure. I divide high schools into five quintiles according to their share of low-income (i.e., SISBEN eligible) students and classify students as *non-low-income* if they attend a high school in the bottom quintile of the distribution of eligibility rates. Students in the next four quintiles are considered *low-income*. Colombia suffers from strong and persistent socioeconomic segregation in secondary education, which makes this school-level proxy of socioeconomic status attractive for the empirical strategy. The (student-weighted) average eligibility rate of high schools in the bottom quintile was 13%, but it more than tripled by the next quintile, reaching 40.3%. For the next three quintiles (from the third to the fifth), the average shares were 58%, 71.1%, and 83.5%.

Since the previous definitions are somewhat arbitrary, I consider alternative definitions of being a low-income student. In particular, the results are robust to using the exact individual-level SISBEN eligibility with data from Laajaj et al. (2022b) for the year 2015 and to using the self-reported proxy for SISBEN eligibility with the main dataset otherwise used throughout the article.

2.4.3 Working Sample, Key Outcome and Summary Statistics

The main working sample consists of students 14-23 years old from years 2014-2018 for whom we have information about their school's SISBEN eligibility rate in 2014. The sample size is approximately 493,000 student-level observations per year. Following Laajaj et al. (2022a), the key outcome used is the student's "global score" on the SABER 11 test, which was the one that mattered in terms of eligibility purposes for SPP. This overall score is a linear combination of scores in different subjects (mathematics, critical reading, natural sciences, social sciences, and English). I follow ICFES and Londoño-Vélez et al. (2020a) to compute the average as follows[31]

$$\text{SABER 11 score}_i = 5 \times \left(\frac{3Math_i + 3Reading_i + 3NatSci_i + 3SocSci_i + Eng_i}{13} \right). \qquad (20)$$

The raw average score ranges from 0 to 500 points, but for the analysis, scores from all years 2014-2018 are standardized using the mean and standard deviation of the 2014 overall distribution. The structure of the SABER 11 exam changed starting with the test administered in the second semester of 2014 (the test taken by the first cohort eligible for SPP). As a consequence, scores are not comparable between the periods 2011-2013 and 2014-2018. Nevertheless, I employ data from 2011-2013 for some supplementary analyses and robustness checks.

Figure D3 shows the distribution of standardized test scores by gender and income groups, as defined in subsubsection 2.4.2, in 2014. Non-low-income students of both genders perform better than their low-income counterparts. However, there is a substantial overlap between the distributions of both groups. For both income groups, the male distribution of test scores exhibits a "higher" right-tail, reflecting the gender gap discussed in subsubsection 2.2.2.

Summary statistics are provided in Table D2. The table shows the average of individual- and school-level characteristics for male and female students in 2014, testing the within-gender differences between students classified as low-income and non-low-income. The table also tests the difference between low-income male students and low-income female students. In general,

[31] Students receive a rounded version of this average score as part of their individual results, which is available in the administrative records from ICFES. I compute it from first principles to avoid this rounding.

the table shows that students classified as non-low-income exhibit a higher degree of characteristics typically associated with better test score performance. They tend to be younger; their school is less likely to be public and located in a rural location; it is also more likely to be on a full-day schedule. Non-low-income students also have more educated parents and smaller families. Overall, this is not surprising and reflects that the school-level SISBEN measure is capturing relatively well the socioeconomic status of the students (and the characteristics of the schools). Moreover, when comparing male students from a low-income background to their female peers, the table shows that men are relatively positively selected, which goes in line with the gendered patterns in dropout behavior discussed before. However, these differences tend to be smaller relative to the ones between students from different socioeconomic statuses.

2.5 Empirical Analysis

2.5.1 Change in the Relative Performance at the Top

Before introducing the main non-linear difference-in-differences strategy in the next section, I first describe a simple analysis in which I investigate the change in the likelihood of performing at the top of the *overall* SABER 11 test score distribution before and after the introduction of SPP for low-income students relative to non-low-income students, separately for male and female students. Note that my focus on the top of the distribution is motivated by the target of the exogenous variation in financial aid that I am exploiting. In particular, I estimate the following equation by Ordinary Least Squares (OLS) for men and women:

$$\text{Top } j_{i,t(i)} = \delta_{0,j} + \delta_{1,j} \left(LowInc_i \times Post_{t(i)} \right) + X_{i,t(i)} \Lambda_{j,t(i)} + \varepsilon_{i,t(i)} \quad \text{for} \quad j \in \{10, 5\}, \tag{21}$$

where Top $j_{i,t(i)}$ is an indicator variable equal to one if student i in test year t performs at the top j percent of the distribution of test scores among all students in their cohort. $Male_i$ is an indicator for male students, $LowInc_i$ is an indicator for low-income students proxied by attending a school where the majority of students are eligible for SPP on the SISBEN margin as explained in subsection 2.4, and $Post_{t(i)}$ is an indicator for years 2015-2018 (compared to 2014). $X_{i,t(i)}$ is a vector of characteristics of the students, including fixed effects for the schools they attended, and indicators for age, education of the mother, education of the father, family size, and the school's municipality (county) location, all interacted with year indicators. Finally, $\varepsilon_{i,t(i)}$ is an error term.

The coefficient of interest is $\delta_{1,j}$, representing the change in the likelihood of performing at the top of the distribution of test scores for low-income students relative to their non-low-income peers.

2.5.2 Non-linear Difference-in-Differences Model: Identification and Estimation Approach

I use a difference-in-differences strategy to evaluate the impact of *Ser Pilo Paga* on the gender achievement gap. Since the scholarship was targeted at low-income students performing at the top of the distribution of the SABER 11 test, I employ a *non-linear* difference-in-differences approach that allows me to study the impact of the policy across the distribution of test scores and not just at the mean. More precisely, I compare the gender achievement gap before and after the introduction of SPP and between low-income and non-low-income students, using the evolution of the entire distribution of non-low-income students to learn about the counterfactual distribution of low-income students. The estimation of the distributional effects of SPP on the gender gap requires the estimation of *two* counterfactual distributions: the counterfactual distribution of low-income men and the counterfactual distribution of low-income women that would have prevailed in the absence of the program. To achieve this, I first perform a gender-specific analysis comparing the test scores of low-income students to their non-low-income peers from the same gender at a given quantile and then calculate the change in the gender gap as the difference in the response of men and women at the same quantile. The high-level assumption behind this strategy is that changes in the distribution of test scores among non-low-income students can be used to identify the counterfactual distribution of low-income students' test scores, allowing me to identify what would have been the observed gender achievement gap in the absence of SPP. More specifically, the non-linear difference-in-differences approach I follow relies on the

Changes-in-Changes model (CiC) proposed by Athey and Imbens (2006). I extend their model to accommodate the fact that I need to estimate a difference involving two counterfactual distributions, as mentioned earlier. In what follows, I first briefly present the CiC model and its assumptions and then describe the particular procedure I conduct to estimate the impact of SPP on the gender achievement gap.

Adapting the notation from Athey and Imbens (2006), let $S_{i,g(i),\ell(i),t(i)}$ denote the SABER 11 test score of student i of gender $G(i) = g \in \{f, m\}$ and income group $L(i) = \ell \in \{0(\text{NonLowInc}), 1(\text{LowInc})\}$, observed in time period $T(i) = t \in \{0(\text{Before SPP}), 1(\text{After SPP})\}$. For simplicity, I will write $S_{i,g(i),\ell(i),t(i)}$ as $S_{g,\ell,t}$ later. Furthermore, let $S^N_{i,g(i)}$ be the test score of student i of gender g in the absence of SPP and $S^I_{i,g(i)}$ be the score for the same student with the program in place.

The CiC model assumes that the outcome in the absence of the intervention satisfies $S^N_{i,g(i)} = h_g(U_i, T_i)$, where $h_g(u, t)$ is a general production function restricted to be strictly increasing in u for all t and that does not depend on the income group L. Following Athey and Imbens (2006), U represents a single index capturing the unobserved characteristics of the student affecting the outcome (e.g., student's ability). Like the traditional mean difference-in-differences model, the CiC model also assumes that the composition of the treatment and comparison groups does not differentially change over time in a way related to the program. In other words, the distribution of U must be constant over time within each income group. However, the model does not impose any restrictions on the distribution of unobservables across income groups, and so these distributions can be arbitrarily different. The CiC model additionally includes a common support assumption requiring overlap of the distribution of test scores among low-income and non-low-income students before and after the policy. This last assumption can be somewhat relaxed by limiting the range of the outcome for which treatment effects can be estimated.

Under the previous assumptions, Athey and Imbens (2006) show that the counterfactual distribution of test scores for low-income students of gender g, $F_{S^N_{g,1,1}}(s)$, can be identified by

$$F_{S^N_{g,1,1}}(s) = F_{S_{g,1,0}}\left(F^{-1}_{S_{g,0,0}}\left(F_{S_{g,0,1}}(s)\right)\right), \tag{22}$$

where $F_{S_{g,\ell,t}}$ and $F^{-1}_{S_{g,\ell,t}}$ denote the cumulative distribution function and inverse cumulative distribution function of $S_{g,\ell,t}$, respectively. For a given gender, the CiC model uses the pre- and post-SPP distributions of non-low-income students to recover the counterfactual distribution of the low-income group. Using the previous result, the effect of SPP on the gender achievement gap at quantile q of the distribution of $F_{S_{m,1,0}}$ and $F_{S_{f,1,0}}$, Diff-QTT$^{CiC}(q)$, will be given by:

$$\begin{aligned}
\text{Diff-QTT}^{CiC}(q) &= F^{-1}_{S^I_{m,1,1}}(q) - F^{-1}_{S^I_{f,1,1}}(q) - \left[F^{-1}_{S^N_{m,1,1}}(q) - F^{-1}_{S^N_{f,1,1}}(q)\right] \\
&= F^{-1}_{S^I_{m,1,1}}(q) - F^{-1}_{S^N_{m,1,1}}(q) - \left[F^{-1}_{S^I_{f,1,1}}(q) - F^{-1}_{S^N_{f,1,1}}(q)\right] \\
&= \text{QTT}^{CiC}_m(q) - \text{QTT}^{CiC}_f(q),
\end{aligned} \tag{23}$$

with $\text{QTT}^{CiC}_m(q)$ and $\text{QTT}^{CiC}_f(q)$ defined as the traditional CiC estimators introduced by Athey and Imbens (2006):

$$\text{QTT}^{CiC}_m(q) = F^{-1}_{S^I_{m,1,1}}(q) - F^{-1}_{S^N_{m,1,1}}(q) = F^{-1}_{S^I_{m,1,1}}(q) - F^{-1}_{S_{m,0,1}}\left(F_{S_{m,0,0}}\left(F^{-1}_{S_{m,1,0}}(q)\right)\right) \tag{24}$$

$$\text{QTT}^{CiC}_f(q) = F^{-1}_{S^I_{f,1,1}}(q) - F^{-1}_{S^N_{f,1,1}}(q) = F^{-1}_{S^I_{f,1,1}}(q) - F^{-1}_{S_{f,0,1}}\left(F_{S_{f,0,0}}\left(F^{-1}_{S_{f,1,0}}(q)\right)\right). \tag{25}$$

For a given gender g, starting with the qth quantile of the pre-SPP distribution of test scores of low-income students (e.g., 290 points for the 90th quantile), the CiC model locates the corresponding quantile for that outcome value in the pre-SPP distribution of the non-low-income group (e.g., 290 points corresponds to the 60th quantile) and then follows the evolution over time of such quantile (e.g., the 60th quantile of the post-SPP distribution of non-low-income students is 291). The *Changes-in-Changes Quantile Treatment Effect on the Treated* (QTT^{CiC}) at quantile q is then the difference between the observed test score for the post-SPP qth quantile among the low-income students (e.g., 293 points for the 90th quantile) minus the counterfactual value derived before (i.e., 291 points), or 2 points (e.g., 0.04 standard deviations if the pre-SPP overall standard deviation is 50 points). The effect of the policy on the gender achievement gap at quantile q, $\text{Diff-QTT}^{CiC}(q)$, will then be given by the difference between the QTT^{CiC} for men and women at quantile q: $\text{QTT}_m^{CiC}(q) - \text{QTT}_f^{CiC}(q)$. The CiC model is attractive because it allows one to construct the gender-specific counterfactual distributions by comparing the evolution of outcome *levels* over time instead of outcome *quantiles*, which I believe is more relevant in the current setting, given the pre-SPP differences in the distributions of test scores between low-income and non-low-income students (see Figure D3). After this, one can compute the change in the gender gap as the difference between two "horizontal" differences.

Empirically, I use the corresponding empirical CDFs, $\widehat{F}_{S_{g,\ell,t}}$, to obtain $\widehat{\text{QTT}}_m^{CiC}(q)$, $\widehat{\text{QTT}}_f^{CiC}(q)$, and $\widehat{\text{Diff-QTT}}^{CiC}(q)$ by using the plug-in principle.[32] For inference, I rely on a non-parametric stratified bootstrap procedure in which I draw an equally-sized random sample with replacement from the original sample of students, stratifying by gender, low-income status, and year. I draw 5,000 bootstrap samples in this way and report 90 and 95 percent confidence intervals.

I calculate $\widehat{\text{Diff-QTT}}^{CiC}(q)$ for pairs of years separately, always using 2014 as the pre-SPP base period. For example, I calculate the effect of SPP on the gender achievement gap for 2016 by using the years 2014 and 2016 only. To summarize the effects of the policy in a single number for each quantile, I proceed in two different ways. Following Havnes and Mogstad (2015), I collapse the data into treatment-by-period groups by pooling all the post-SPP years together. Alternatively, I also compute a weighted average of year-specific effects using the sample size of

[32] I create my own program using Kranker (2019)'s Stata command as a base, which, in turn, is based on the original Matlab code by Athey and Imbens (2006).

each individual year as weights. I prefer this last approach since aggregating after comparing students within the same cohort seems more appropriate.

2.6 Results

2.6.1 Results for Changes in the Relative Performance at the Top

Table 6 shows estimates of $\delta_{1,j}$ from Equation 21 measuring the change in the probability of performing at the top of the *overall* distribution of test scores for low-income students relative to non-low-income students, separately by gender. The share of low-income men performing at the top before SPP is almost twice as big as the share of low-income women. This gap also exists among non-low-income students but is relatively smaller. Table 6 reveals a negative but statistically insignificant change in this probability among low-income men for the top 10% and an increase of 0.53 percentage points (p.p.) for the top 5%, a jump of almost 16% from the 2014 mean. For women, these changes are 0.47 p.p. (up 10% from the 2014 mean) and 0.61 p.p. (35.6%), respectively. Figure D4 adds additional evidence that the relative changes in these probabilities are not significantly driven by pre-trends. Overall, these results suggest that the program had a stronger effect on female students' performance.

Table 6. Change in the Probability of Scoring at the Top of the SABER 11 Distribution

	Top 10%		Top 5%	
	Male	Female	Male	Female
	(1)	(2)	(3)	(4)
$LowInc \times Post$ (×100)	-0.234	0.473*	0.527**	0.606***
	(0.291)	(0.251)	(0.235)	(0.194)
Observations	1,124,456	1,341,173	1,124,456	1,341,173
Pre-SPP low-income mean (%)	8.0	4.7	3.4	1.7
Pre-SPP non-low-income mean (%)	34.8	29.3	22.1	16.8
Change in low-income mean (%)	-2.9	10.1	15.6	35.6

Notes: This table presents estimates from Equation 21. Each column presents estimates from a linear probability model with scoring at the top j percent of the *overall* distribution of SABER 11 as the outcome, with $j \in \{10, 5\}$. The regressions include school fixed effects and indicators for age, mother's education, father's education, family size, and municipality (county) of the school (all interacted with year indicators). The estimated coefficients have been multiplied by 100, so they should be directly interpreted as percentage points. Robust standard errors are shown in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

2.6.2 Non-linear Difference-in-Differences Results

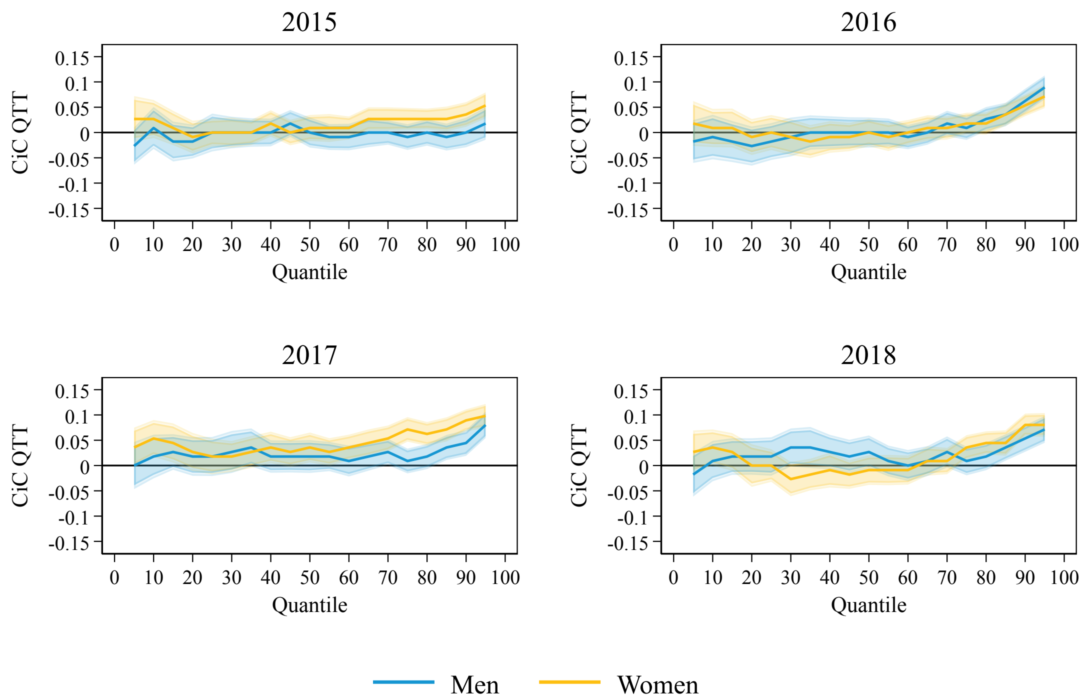
I start by showing estimates for $\text{QTT}_g^{CiC}(q)$ from Equation 24 and Equation 25 for $q \in \{5, 10, \ldots, 95\}$. As documented by Laajaj et al. (2022a) for the year 2015, Figure 7 shows an increase in performance on the SABER 11 test concentrated at the top of the distribution for the four cohorts following the announcement of SPP (2015-2018). For both low-income men and low-income women, there is a significant increase in test scores starting around the 60th-70th percentiles of the distribution, with bigger estimated increases after 2015, which is consistent with the dynamics of the aggregate results in Laajaj et al. (2022a) for SABER 9 test scores among younger students.

However, the plot shows that the increase in test scores at the top is, in general, stronger for female students, in particular during the years 2015 and 2017. In Figure 8, I formally test this differential response by showing estimates for Diff-$\text{QTT}^{CiC}(q)$ from Equation 23. In Table 7, I present the yearly and post-SPP summary results of these estimates for selected quantiles. On average, I find a 0.033 standard deviations reduction (16.2% from the 2014 mean) in the gender achievement gap at the 75th percentile of the test scores distribution and a 0.024 standard deviations reduction (9.1% from the 2014 mean) at the 90th percentile.[33] These average results mask some heterogeneity over the years. They are mainly driven by the significant differential response in the years 2015 and 2017 (and, to some extent, 2018). Unfortunately, the model presented in subsection 2.3 does not provide straightforward implications to explain these patterns in *differential* responses over time, and the estimated effects do not exhibit any clear dynamic pattern.

The only significant change in the gender gap detected at lower quantiles is found around the 30th quantile in the year 2018. As shown in Figure 7, this is driven by the biggest differential increase in test scores for men observed in that year when compared to their relative performance

[33] These results correspond to weighted averages over the four years following the announcement of SPP. Table D4 shows results pooling all the post-SPP cohorts together. The results are similar.

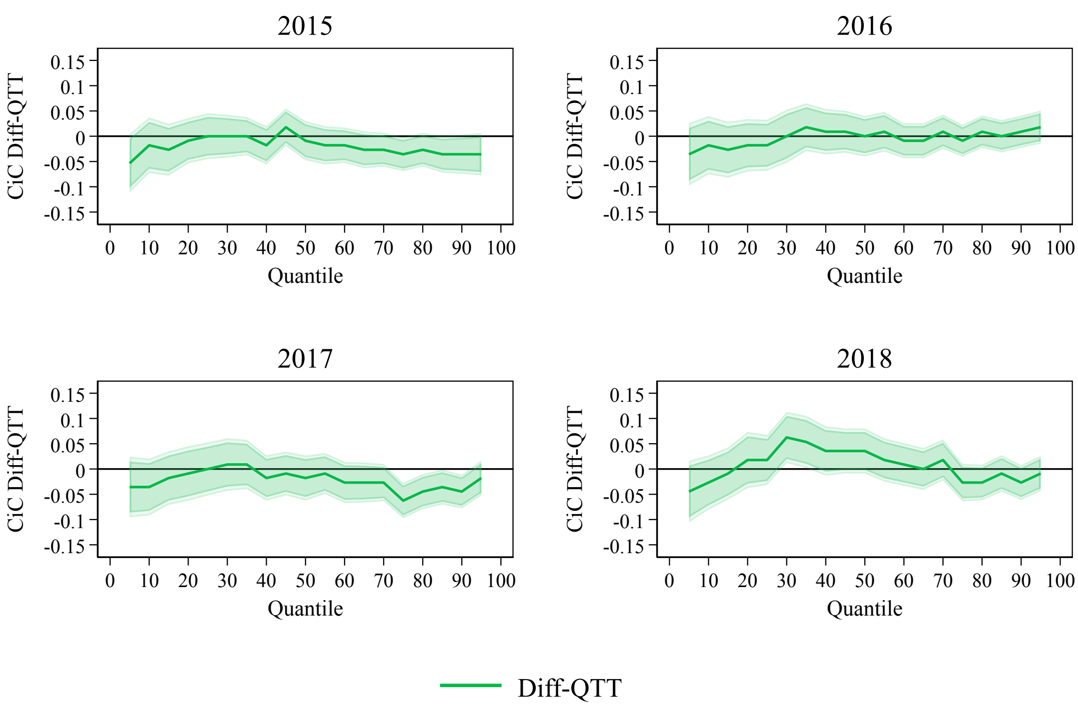
Figure 7. Distributional Effects of *Ser Pilo Paga* on SABER 11 Test Scores by Gender



Notes: This figure presents estimates from Equation 24 and Equation 25 for years 2015-2018. For each quantile $q \in \{5, 10, \dots, 95\}$, the solid lines report $\widehat{\text{QTT}}_m^{CiC}(q)$ and $\widehat{\text{QTT}}_f^{CiC}(q)$. The darkest shaded area represents 90 percent confidence intervals, while the lightest represents 95 percent confidence intervals. Confidence intervals are based on 5,000 bootstrap replications stratifying by low-income status and year (for a given gender).

during the 2015-2018 period. Beyond this, the most significant changes in the gap are in favor of women and concentrated at the top of the distribution.

Figure 8. Distributional Effects of *Ser Pilo Paga* on the Gender Achievement Gap



Notes: This figure presents estimates from Equation 23 for years 2015-2018. For each quantile $q \in \{5, 10, \ldots, 95\}$, the solid lines report $\widehat{\text{Diff-QTT}}^{CiC}(q)$. The darkest shaded area represents 90 percent confidence intervals, while the lightest represents 95 percent confidence intervals. Confidence intervals are based on 5,000 bootstrap replications stratifying by gender, low-income status, and year.

Table 7. Non-linear Difference-in-Differences Estimates of the Effects of *Ser Pilo Paga* on the Gender Achievement Gap

	2015	2016	2017	2018	Average	Initial gap	Average change (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mean	-0.020	-0.004	-0.023	0.006	-0.010	0.134	-7.6
90% CI	[-0.041,0.002]	[-0.027,0.020]	[-0.047,0.002]	[-0.020,0.031]	[-0.029,0.009]		
95% CI	[-0.046,0.006]	[-0.032,0.024]	[-0.052,0.007]	[-0.025,0.036]	[-0.033,0.012]		
Diff-QTT(25)	0.000	-0.018	0.000	0.018	0.000	0.054	0.0
90% CI	[-0.038,0.038]	[-0.060,0.024]	[-0.044,0.044]	[-0.024,0.059]	[-0.033,0.033]		
95% CI	[-0.045,0.045]	[-0.068,0.032]	[-0.053,0.053]	[-0.032,0.067]	[-0.039,0.039]		
Diff-QTT(50)	-0.009	0.000	-0.018	0.036	0.002	0.134	1.6
90% CI	[-0.041,0.024]	[-0.034,0.034]	[-0.055,0.020]	[-0.002,0.073]	[-0.025,0.029]		
95% CI	[-0.048,0.030]	[-0.040,0.040]	[-0.062,0.027]	[-0.009,0.080]	[-0.030,0.034]		
Diff-QTT(75)	-0.036	-0.009	-0.062	-0.027	-0.033	0.205	-16.2
90% CI	[-0.064,-0.007]	[-0.035,0.018]	[-0.092,-0.033]	[-0.058,0.004]	[-0.056,-0.011]		
95% CI	[-0.069,-0.002]	[-0.040,0.023]	[-0.098,-0.027]	[-0.064,0.010]	[-0.060,-0.006]		
Diff-QTT(90)	-0.036	0.009	-0.045	-0.027	-0.024	0.268	-9.1
90% CI	[-0.068,-0.003]	[-0.018,0.036]	[-0.073,-0.017]	[-0.056,0.002]	[-0.047,-0.002]		
95% CI	[-0.074,0.003]	[-0.023,0.041]	[-0.078,-0.011]	[-0.061,0.008]	[-0.051,0.003]		
Observations (2014)	504,171	504,171	504,171	504,171	504,171		
Observations (t)	496,949	499,562	484,347	480,600	1,961,458		

Notes: This table presents estimates from Equation 23 for selected quantiles and the mean CiC effect. The reference period ($t = 0$) for all columns is 2014. The post-SPP average estimate reported in column (5) is the weighted average of the year-specific effects for 2015-2018 using the number of observations from each year as weights. Confidence intervals (CI) are based on 5,000 bootstrap replications stratifying by gender, low-income status, and year.

I use the results in Table 7 to compare the estimated reduction in the gender achievement gap with the reduction in the *socioeconomic* gap documented by Laajaj et al. (2022a). The estimated reductions in the socioeconomic gap between SISBEN eligible and non-eligible students in SABER 11 scores were 0.059 (11.6% from the 2013-2014 mean) and 0.070 (9.9%) standard deviations at the 75th and 90th percentiles, respectively. The average estimated reductions in the gender achievement gap represent 55.9% (0.033/0.059) and 34.3% (0.024/0.070) of these reductions.[34]

One potential concern is that the estimated reductions in the gender achievement gap presented in Table 7 are due to differential changes in the characteristics of men and women taking the SABER 11 test over time. If SPP differentially induced students from a given gender to stay in high school, then the estimates are contaminated by sample compositional changes. However, if the patterns shown in Figure 4 are somewhat indicative of the type of selection in test taking, the selection affecting the gender gap is most likely not coming from the top part of the distribution of test scores.

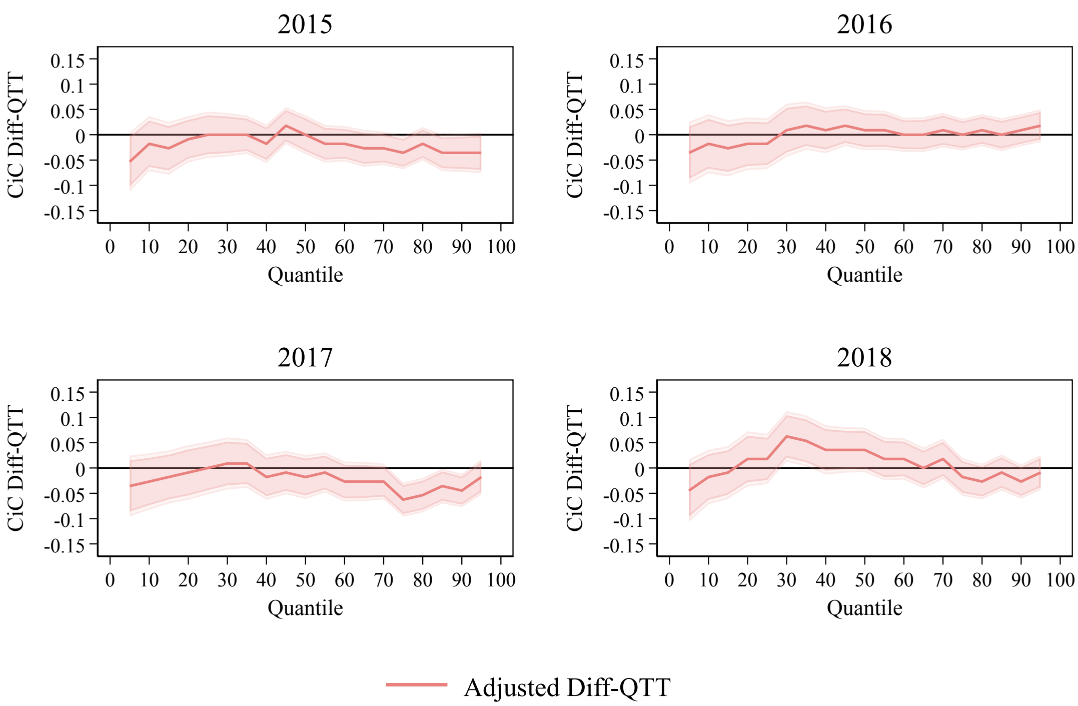
Nevertheless, I use a re-weighting procedure in the spirit of DiNardo, Fortin and Lemieux (1996) to control for these potential compositional changes that involves re-weighting the empirical CDFs of low-income students in the post-SPP years. Under the assumption that the group of non-low-income students was not affected by these potential changes, this procedure should isolate the effect of the program from the effect of the sample composition. The fundamental assumption underlying the exercise is, however, that the weights being used are valid to account for these changes (that can potentially come from unobserved characteristics). In particular, I estimate the following weights parametrically using a probit model:

$$\begin{aligned}\psi_{i,\tau} = {} & Male_i \times \frac{\Pr(Male = 1|X, t = 2014)}{\Pr(Male = 1|X, t = \tau)} \\ & + (1 - Male_i) \times \frac{1 - \Pr(Male = 1|X, t = \tau)}{1 - \Pr(Male = 1|X, t = 2014)},\end{aligned} \tag{26}$$

[34]Note, however, that the results in Laajaj et al. (2022a) come from a variation of a regression discontinuity design and reflect local effects and, therefore, are not directly comparable to my estimates. Additionally, Laajaj et al. (2022a) only consider one year post-SPP in their estimations (2015).

for $t = \tau \in \{2015, 2016, 2017, 2018\}$. The vector X includes indicators for age, mother's education, father's education, family size, public school, rural school, and school schedule. These are the same covariates used to construct the weights in Equation 6. For inference, these weights are re-estimated each time during the bootstrap procedure. Results are presented in Figure 9 and Table 8. In general, the estimated reductions are almost identical or slightly smaller in absolute value. The average reduction in the gap at the 75th percentile goes from 0.033 to 0.029 standard deviations (12.9% from the adjusted gap in 2014). At the 90th percentile, the estimated reduction in the gap remains unchanged at 0.024 standard deviations (9.1% from the adjusted gap in 2014). These results represent 49.2% (0.029/0.059) and 34.3% (0.024/0.070) of the estimated reductions in the socioeconomic achievement gap reported by Laajaj et al. (2022a) for the same percentiles. Figure D5 presents the average reduction in the gap with and without adjustment for changes in covariates in the same plot to facilitate the comparison.

Figure 9. Distributional Effects of *Ser Pilo Paga* on the Gender Achievement Gap Adjusting by Potential Compositional Changes.



Notes: This figure presents estimates from Equation 23 for years 2015-2018. For each quantile $q \in \{5, 10, \ldots, 95\}$, the solid lines report $\widehat{\text{Diff-QTT}}^{CiC}(q)$ reweighting the post-SPP distributions using the weights discussed in the text. The darkest shaded area represents 90 percent confidence intervals, while the lightest represents 95 percent confidence intervals. Confidence intervals are based on 5,000 bootstrap replications stratifying by gender, low-income status, and year.

Table 8. Non-linear Difference-in-Differences Estimates of the Effects of *Ser Pilo Paga* on the Gender Achievement Gap (Adjusted)

	2015	2016	2017	2018	Average	Initial gap	Average change (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mean	-0.018	0.000	-0.022	0.008	-0.008	0.149	-5.7
90% CI	[-0.040,0.003]	[-0.024,0.023]	[-0.047,0.002]	[-0.017,0.033]	[-0.027,0.010]		
95% CI	[-0.044,0.007]	[-0.028,0.027]	[-0.051,0.006]	[-0.022,0.038]	[-0.030,0.013]		
Diff-QTT(25)	0.000	-0.018	0.000	0.018	0.000	0.071	0.0
90% CI	[-0.038,0.038]	[-0.060,0.024]	[-0.044,0.044]	[-0.024,0.060]	[-0.032,0.032]		
95% CI	[-0.046,0.046]	[-0.068,0.032]	[-0.052,0.052]	[-0.032,0.068]	[-0.039,0.038]		
Diff-QTT(50)	0.000	0.009	-0.018	0.036	0.007	0.161	4.1
90% CI	[-0.032,0.032]	[-0.024,0.042]	[-0.054,0.018]	[-0.001,0.073]	[-0.020,0.033]		
95% CI	[-0.038,0.038]	[-0.031,0.049]	[-0.061,0.025]	[-0.008,0.080]	[-0.025,0.039]		
Diff-QTT(75)	-0.036	0.000	-0.062	-0.018	-0.029	0.223	-12.9
90% CI	[-0.063,-0.008]	[-0.026,0.026]	[-0.091,-0.034]	[-0.049,0.013]	[-0.051,-0.007]		
95% CI	[-0.068,-0.003]	[-0.031,0.031]	[-0.096,-0.029]	[-0.055,0.019]	[-0.055,-0.003]		
Diff-QTT(90)	-0.036	0.009	-0.045	-0.027	-0.024	0.268	-9.1
90% CI	[-0.067,-0.004]	[-0.018,0.036]	[-0.072,-0.017]	[-0.055,0.002]	[-0.047,-0.002]		
95% CI	[-0.073,0.002]	[-0.023,0.041]	[-0.078,-0.012]	[-0.061,0.007]	[-0.051,0.002]		
Observations (2014)	504,171	504,171	504,171	504,171	504,171		
Observations (t)	496,949	499,562	484,347	480,600	1,961,458		

Notes: This table presents estimates from Equation 23 for selected quantiles and the mean CiC effect. The reference period ($t = 0$) for all columns is 2014. The post-SPP average estimate reported in column (5) is the weighted average of the year-specific effects for 2015-2018 using the number of observations from each year as weights. The initial gap displayed in column 6 is the adjusted gap. Confidence intervals (CI) are based on 5,000 bootstrap replications stratifying by gender, low-income status, and year.

2.7 Robustness

2.7.1 Falsification Tests

Using SABER 11 data from 2011 to 2014 (before the announcement of SPP), I provide evidence in Figure D6 that the gender achievement gap was not trending in a particular direction over time before SPP. The figure presents results from a placebo-in-time strategy using Equation 23 and data from 2011-2014 with 2011 as the reference period, assuming that SPP was introduced in each of these years instead of 2014.[35] There are no statistically significant coefficients at the top of the distribution for 2012 and 2013. For 2014, there is a significant and *positive* estimate for the 75th percentile. I note, however, that the SABER 11 exam changed its structure starting in 2014, and therefore the test scores are not directly comparable between 2011 and 2014. For the two years that are fully comparable with 2011 (2012 and 2013), the change in the gap at the top is relatively flat. This exercise, together with the results in Figure D4, supports that the results are not driven by time trends in the gender gap predating SPP.

2.7.2 Alternative Treatment Definitions

As mentioned earlier, while useful, the definitions of low-income and non-low-income students used in the main analysis are somewhat arbitrary. Because of this, I report in Figure D7 and Table D6 that the results are robust to using alternative definitions of these groups. In particular, Panel A of Figure D7 shows the estimated effects of the reduction in the gender achievement gap for 2015 (the only post-SPP available in the data) using true individual-level SISBEN eligibility for SPP coming from Laajaj et al. (2022b). I use data for 2014 and 2015 from Laajaj et al. (2022b) and apply the same sample restrictions as in the main analysis (i.e., keeping students 14-23 years old) and then estimate Equation 23. I estimate a statistically significant

[35] For years 2011-2013, I use the mean and standard deviation of the overall distribution in 2011 to standardize the scores. Following Londoño-Vélez et al. (2020a), the global average for these years is calculated as

$$\text{SABER 11 score}_i = \frac{Chem_i + Bio_i + Phys_i + 2SocSci_i + Philo_i + 3Lang_i + 3Math_i + Eng_i}{13}.$$

reduction in the gender achievement gap of 0.047 standard deviations at the 90th percentile using their data, compared to the 0.036 of the main analysis.[36]

On top of this, I also calculate the reduction in the gender gap within the same dataset of the main analysis using a proxy for SISBEN eligibility that is self-reported by students. Students reported their SISBEN level; having a SISBEN level of 1 or 2 was approximately equivalent to meeting the eligibility requirements for SPP on the SISBEN scale. Students with higher SISBEN levels or not categorized were typically ineligible. Panel B of Figure D7 presents these results. Even though the estimated reduction at the 90th percentile is not statistically significant, the estimates for the 80th, 85th, and 95th are. For this last percentile, the estimated reduction is 0.036 standard deviations, which coincides with the one in the main analysis.

2.8 Conclusion and Discussion

In this paper, I showed evidence that limited college-going opportunities have a role in explaining part of the gender achievement gap in standardized tests among high school students. Using a substantial expansion of financial aid for high-performing, low-income students to attend college in Colombia, I presented evidence of this policy reducing the gender achievement gap at the top by about 13% at the 75th percentile and 9% at the 90th percentile of the distribution of test scores.

This paper contributes to the extensive literature aiming to explain the underrepresentation of women at top performance levels in standardized tests by highlighting that limited college-going opportunities likely disproportionally affect women's outcomes. Broadly speaking, it complements the results from Breda et al. (2018), who argue that general societal inequality (and not only gender inequality) affects the performance of female students. I document a particular dimension of inequality: income-driven gaps in access to college. *Ser Pilo Paga* represented a positive shock to the economic opportunities of low-income students, which ended up reducing the gender gap in performance.

[36] The results coming from the Laajaj et al. (2022b)'s dataset are less smooth than the main results, in part because their dataset has a rounded version of the raw SABER 11 score for the years 2014 and 2015.

The estimated decrease in the gender achievement gap at the top is meaningful when compared to the overall reduction in the *socioeconomic* gap between low-income and non-low-income students documented by Laajaj et al. (2022a). The findings in the paper are also in line with the results of Londoño-Vélez et al. (2023) showing that the program had a greater positive impact on female students compared to male students in terms of increased opportunities to enroll in and successfully complete undergraduate programs at high-quality universities. Furthermore, I posit that the differential response of women that is behind the reduction in the gender gap mattered for the allocation of scholarships. Using a simple back-of-the-envelope calculation in which I use the counterfactual distribution of test scores for low-income women that would have prevailed if their effort response was the same as the one for men at each quantile and the evolution of the SABER 11 threshold for SPP over time, I estimate that women's differential response translated into roughly 900 additional low-income women becoming SABER 11 eligible for SPP between 2015 and 2017, which represents around 7% of the number of scholarships granted by the SPP program to female students during that period.

The gender gap in standardized test scores is still high in Colombia, particularly among top-performing students. Since these tests are crucial in determining financial aid and access to high-quality post-secondary education, further research is needed to evaluate how to keep closing the gender gap in performance.

Appendix A. Additional Tables and Figures for Chapter 1

Table A1. Municipality Characteristics

	Comparison municipality mean (1)	Treatment municipality difference (2)
Ln(population) (2008)	9.379	0.402***
	(0.040)	(0.064)
Rural share of population (2008)	0.617	-0.085***
	(0.010)	(0.014)
Distance to department's capital (km)	87.335	-19.266***
	(2.666)	(3.419)
Distance to nearest SPP eligible institution (km)	111.270	-25.403***
	(4.707)	(6.018)
Poverty incidence (2005)	0.548	-0.066***
	(0.005)	(0.006)
Gini coefficient (0-1) (2005)	0.462	-0.016***
	(0.002)	(0.002)
Public expenditure per capita (2008)	1,170.896	-272.104***
	(179.473)	(180.781)
Tax revenue per capita (2008)	94.599	48.854***
	(4.230)	(8.575)
Public investment in education per capita (2008)	329.604	-109.323***
	(159.372)	(195.344)
Gross enrollment rate 6th-9th grade (2011)	0.975	0.091***
	(0.012)	(0.016)
Gross enrollment rate 10th-11th grade (2011)	0.646	0.140***
	(0.011)	(0.015)
Dropout rate 6th-9th grade, public schools (2011)	0.050	0.002***
	(0.001)	(0.002)
Dropout rate 10th-11th grade, public schools (2011)	0.040	-0.001***
	(0.001)	(0.002)
Exposed to FARC (0/1) (2011-2014)	0.142	-0.042**
	(0.015)	(0.020)
Number of municipalities	526	541

Notes: This table compares pre-SPP characteristics between treatment and comparison municipalities. Columns 1 and 2 present results of a regression of a municipality characteristic on an indicator for being a treatment municipality. Column 1 shows the coefficients on the intercept term and represents the mean of comparison municipalities. Column 2 shows coefficients on the treatment indicator and represents the mean difference between treatment and comparison municipalities. Money variables are measured in 2019 thousand Colombian pesos. Robust standard errors are presented in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$). Significance stars are suppressed for coefficients on the intercept term.

Table A2. Summary Difference-in-Differences Estimates for Teen Fatherhood (Poisson Model)

	Births per 1,000 men			
Age group:	All teens (15-19)	15-17	18	19
	(1)	(2)	(3)	(4)
$SPP \times Post$	-0.081***	-0.076**	-0.112***	-0.066***
	(0.021)	(0.031)	(0.025)	(0.021)
Observations	12,804	12,744	12,792	12,804
Treatment municipalities	541	541	541	541
Comparison municipalities	526	526	526	526
Pre-trends testing p-value	0.398	0.259	0.554	0.200
Pre-SPP share of teen fathers	100	25.7	33.6	40.7

Notes: This table presents summary difference-in-differences estimates following Equation 3 and a Poisson model with the number of births per 1,000 men in each age group as the outcome. For estimation, we use Correia, Guimarães and Zylkin (2020)'s Stata command. Column 1 presents the estimate for all male teens (15-19 years old). Column 2 shows the results for male teens 15-17 years of age. Columns 3 and 4 present the estimates for men 18 and 19 years of age, respectively. All estimates are weighted by the annual population of men in each municipality and age group. Standard errors are clustered at the municipality level and presented in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

Table A3. Robustness to Alternative Specifications and Sample of Municipalities

	Log fertility rate			IHS fertility rate	Fertility rate	
	(1)	(2)	(3)	(4)	(5)	(6)
$Teen \times SPP \times Post$	-0.057***	-0.057***	-0.035***	-0.059***	-0.060***	-3.204***
	(0.011)	(0.014)	(0.008)	(0.011)	(0.011)	(0.701)
						[-4.5%]
Treatment measure	Discrete	Discrete	Continuous	Discrete	Discrete	Discrete
Exclude capital cities	No	Yes	No	No	No	No
Model	Linear	Linear	Linear	Linear	Poisson	Linear
Observations	25,608	24,840	25,608	26,520	26,514	26,520
Treatment municipalities	541	514	–	552	552	552
Comparison municipalities	526	521	–	553	553	553
Pre-trends test p-value	0.985	0.951	0.684	0.946	0.984	0.972

Notes: This table reports results from variations of our main specification and sample of municipalities. Column 1 replicates our main results. In column 2, we exclude the group of 32 municipalities corresponding to capitals of departments. In column 3, we replace SPP^*_m in Equation 4 by the standardized rate of female SABER 11 test takers in 2014 who are eligible for SPP in each municipality ($StdSPP_m$). Our pre-trends test cannot be applied with the continuous version of the treatment in column 3. Following Muralidharan and Prakash (2017), we replace it by a differential linear pre-trends test using pre-SPP data. Specifically, the reported p-values in column 3 comes from testing the significance of β in the equation $Y_{amt} = \alpha + \beta\,(Teen_a \times StdSPP_m \times t) + \gamma_{am} + \gamma_{mt} + \gamma_{ad(m)t} + \epsilon_{amt}$. Column 4 uses the inverse hyberbolic sine (IHS) transformation of the birth rate as the outcome instead of the log transformation. It includes the group of municipalities with zero births in any given period for any age group (38 out of 1,105, or 3.4%). Columns 5 and 6 use the fertility rate (in levels) as the outcome variable instead of the log fertility rate and include all municipalities. Column 5 estimates the model using a Poisson pseudomaximum likelihood (PPML) regression using Correia, Guimarães and Zylkin (2020)'s Stata command. For column 6, the implied percentage change with respect to the pre-SPP mean adolescent fertility rate of treatment municipalities is shown in brackets. All estimates are weighted by the annual population of women in each municipality and age group. Standard errors are clustered at the municipality level and are reported in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

Table A4. Robustness to Alternative Treatment Definitions

	Log fertility rate		IHS births
	(1)	(2)	(3)
Triple difference coefficient	-0.057***	-0.096***	-0.107***
	(0.011)	(0.015)	(0.022)
Across-municipality treatment	Above median SPP eligibility rate	Below median distance SPP-eligible HEI	Above median SPP eligibility rate
Within-municipality comparison	25-29	25-29	15-19, < 8th grade
Sample (age groups)	15-19 & 25-29	15-19 & 25-29	15-19
Observations	25,608	25,608	25,608
Treatment municipalities	541	535	541
Comparison municipalities	526	532	526
Pre-trends test p-value	0.985	0.023	0.369

Notes: This table reports results from variations of our across-municipality treatment definition and within-municipality comparison group. Column 1 replicates our main triple difference results. In column 2, we use the municipality-level distance to the nearest municipality with an SPP-eligible HEI as our across-municipality treatment definition. Treatment municipalities are those below the median distance (closer), while comparison municipalities are those above the median (farther away). In column 3, we use the across-municipality treatment definition as column 1 but replace the 25-29 years old women as the within-municipality comparison group with teenagers with a level of education less than 8th grade. The outcome in column 3 is the inverse hyberbolic sine (IHS) transformation of the number of births from teenagers in each education-municipality-year cell, because some municipalities in our main sample have zero births for certain education groups. All estimates are weighted by the annual population of women in each municipality and age group. Standard errors are clustered at the municipality level and are reported in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

Table A5. Robustness to Alternative Estimators

	Log fertility rate		
Estimation:	OLS	Imputation Estimator	DID
	(1)	(2)	(3)
$Teen \times SPP \times Post$	-0.057***	-0.083***	-0.074***
	(0.011)	(0.015)	(0.024)
Observations	25,608	25,608	5,335
Treatment municipalities	541	541	541
Comparison municipalities	526	526	526
Pre-trends test p-value	0.985	0.985	0.875

Notes: This table reports results using Borusyak et al. (2021)'s imputation estimator and de Chaisemartin and D'Haultfœuille (2022a)'s DID estimator. Column 1 replicates our main triple difference result using OLS to estimate Equation 4. Column 2 reports the treatment effect using the imputation estimator including the same set of fixed effects from Equation 4. See Borusyak et al. (2021) for details. Column 3 reports the average effect from the DID estimator using the same specifications from Equation 4, but we recast the triple difference as a simple difference-in-differences by substracting the within-municipality log fertility rate of older women from the the log fertility rate of teenagers. The department-specific year trends in column 3 are handled non-parametrically. The pre-trends test in column 3 corresponds to the p-value of a joint test of significance for four placebos before the policy introduction. See de Chaisemartin and D'Haultfœuille (2022a) for details. All estimates are weighted by the annual population of women in each municipality and age group. Standard errors clustered at the municipality level are reported in parentheses and for the DID estimator they are based on 5,000 bootstrap replications (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

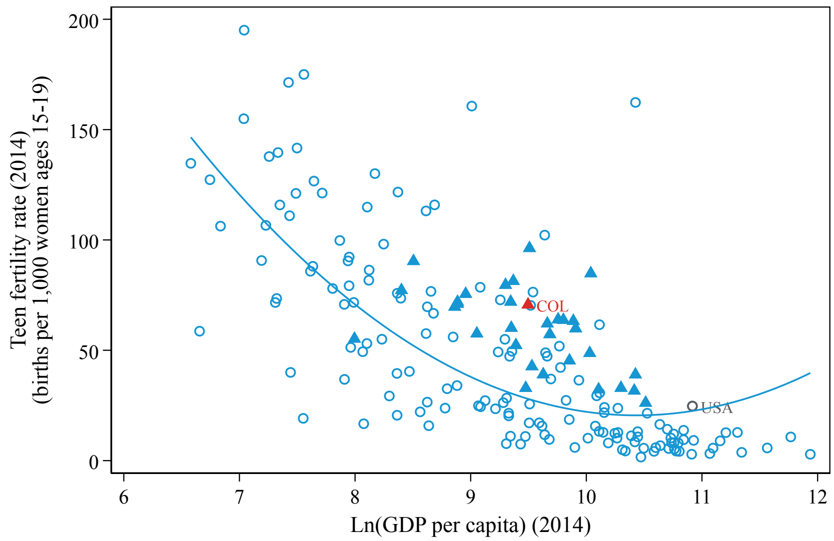
Table A6. Robustness to Placebo Treatment Years

	Log fertility rate					
Placebo year (h):	2008	2009	2010	2011	2012	2013
	(1)	(2)	(3)	(4)	(5)	(6)
$Teen \times SPP \times Post$	-0.007	-0.005	-0.002	-0.001	0.004	0.003
	(0.012)	(0.011)	(0.010)	(0.010)	(0.011)	(0.013)
Observations	14,938	14,938	14,938	14,938	14,938	14,938
Treatment municipalities	541	541	541	541	541	541
Comparison municipalities	526	526	526	526	526	526

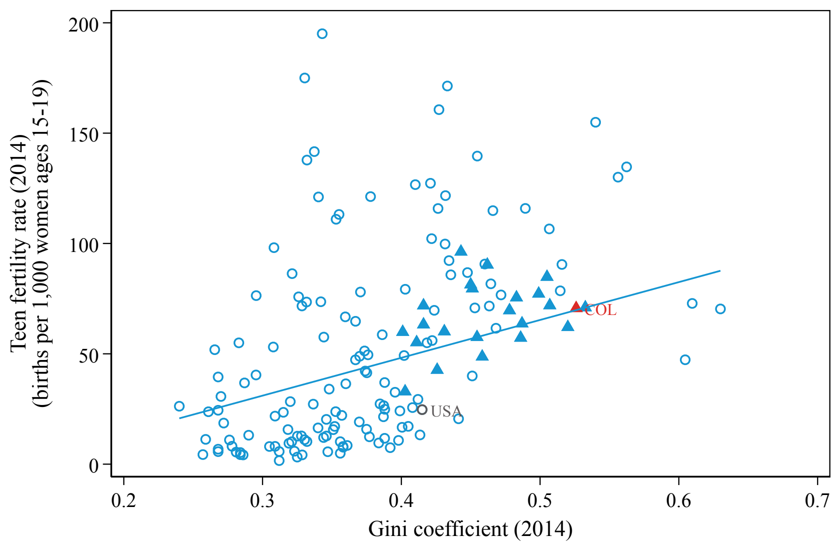
Notes: Each column in this table assumes that SPP was introduced in year h instead of 2014 and estimates Equation 4 with $Post_t = \mathbb{1}[t > h]$ and $t \in [2008, 2014]$. All estimates are weighted by the annual population of women in each municipality and age group. Standard errors are clustered at the municipality level and are reported in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

Figure A1. Correlates of Teen Fertility Rates Across Countries

(a) Teen Fertility Rate and Income per Capita



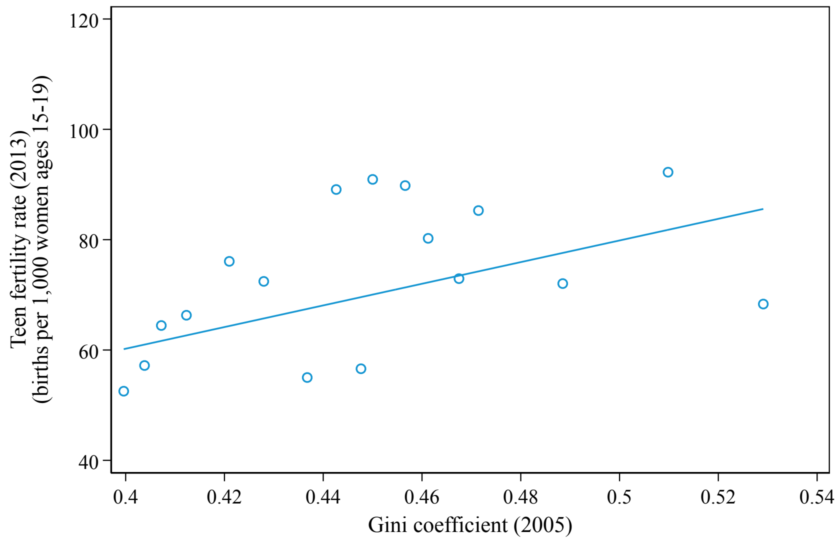
(b) Teen Fertility Rate and Income Inequality



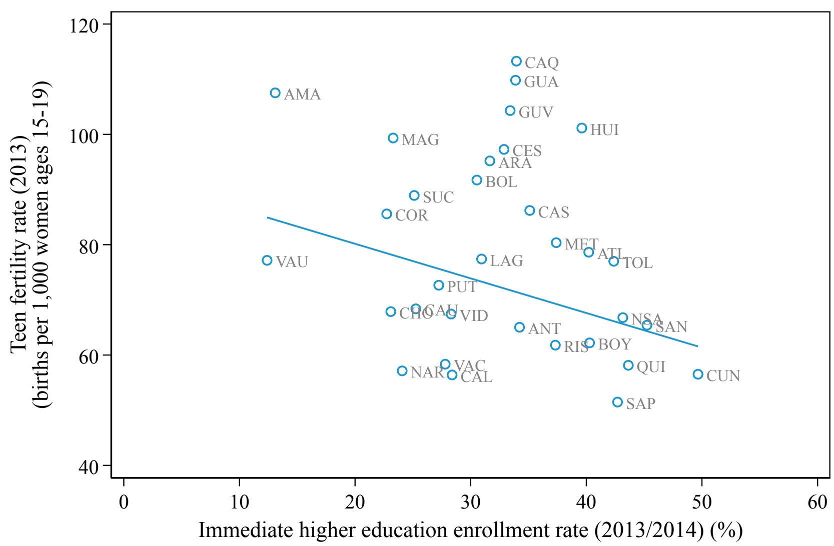
Notes: This figure shows some correlates of teen fertility across countries. Panel (a) shows the relationship between teen fertility rates and GDP per capita (PPP) in 2014 for a sample of 183 countries. Panel (b) shows the relationship between teen fertility rates and the Gini coefficient in 2014 for a sample of 161 countries. In Panel (b), for some countries, the Gini corresponds to the closest year before 2014 in case 2014 was unavailable. The data come from the World Bank's World Development Indicators. Countries in Latin America and the Caribbean are shown in solid triangles, while all other countries are shown in hollow circles. Colombia (COL) is highlighted in red, and the US (USA) position is included for reference. In Panel (a), the solid line corresponds to a quadratic fit weighted by the population of women ages 15-19 in each country. The coefficients are -177.032 (s.e. = 63.185) and 8.484 (s.e. = 3.549). In Panel (b), the solid line corresponds to a linear fit weighted by the population of women ages 15-19 in each country. The slope is 171.453 (s.e. = 57.503).

Figure A2. Correlates of Teen Fertility Rates in Colombia

(a) Teen fertility rate and income inequality across municipalities.

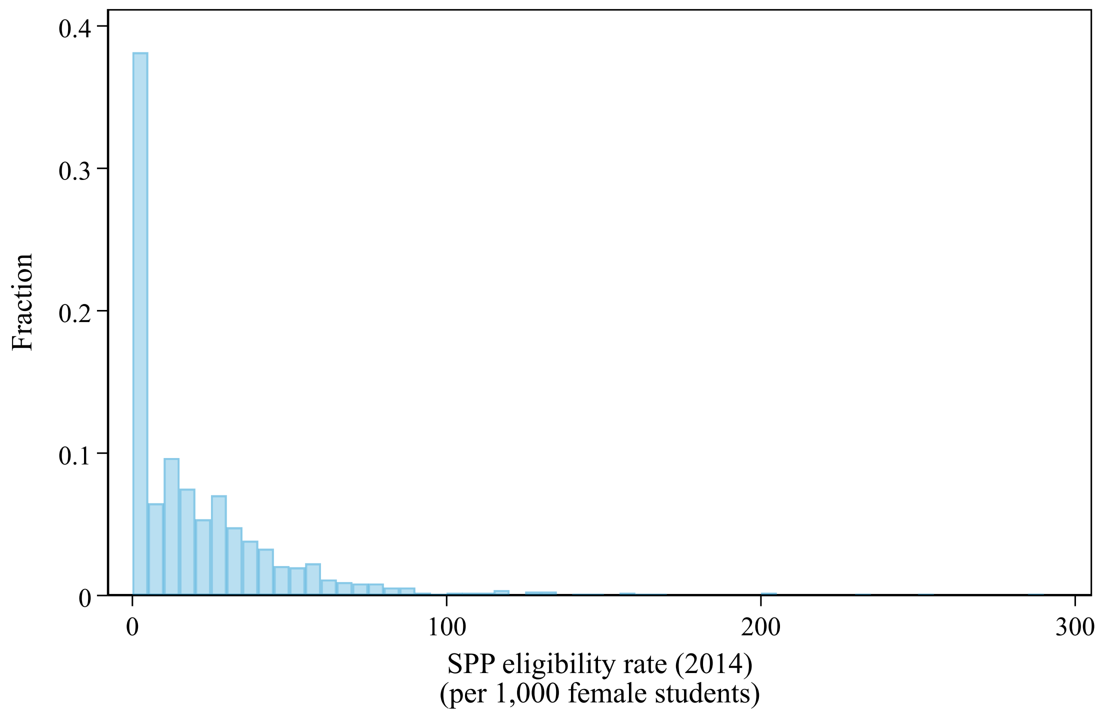


(b) Teen fertility rate and college enrollment across departments.



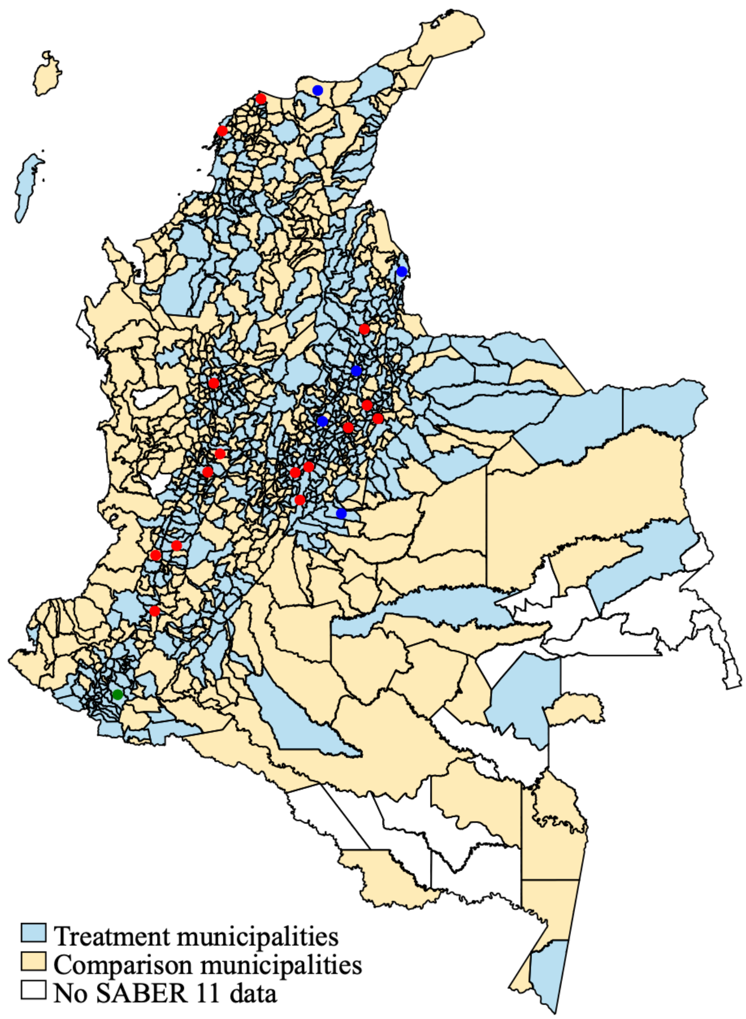
Notes: This figure shows some correlates of teen fertility across municipalities and departments in Colombia. Departments in Colombia are similar to states in the United States. A group of municipalities forms each department. Panel (a) shows a binned scatterplot of the relationship between teen fertility rates in 2013 and the Gini coefficient in 2005 for a sample of 1,040 municipalities. The solid line corresponds to a linear fit weighted by the population of women ages 15-19 in each municipality. The slope is 196.207 (s.e. = 44.515). Panel (b) shows the relationship between teen fertility rates in 2013 and the immediate higher education enrollment rate for the 2013 cohort of high school seniors for the 32 departments in Colombia. This rate is measured as the percentage of students in 11th grade in 2013 that enrolled in higher education in 2014. The solid line corresponds to a linear fit weighted by the population of women ages 15-19 in each department. The slope is -0.627 (s.e. = 0.304). Both panels adjust birth rates for the lag between conception and birth. The data come from the official birth records and the Ministry of Education.

Figure A3. Distribution of SPP Eligibility Rates



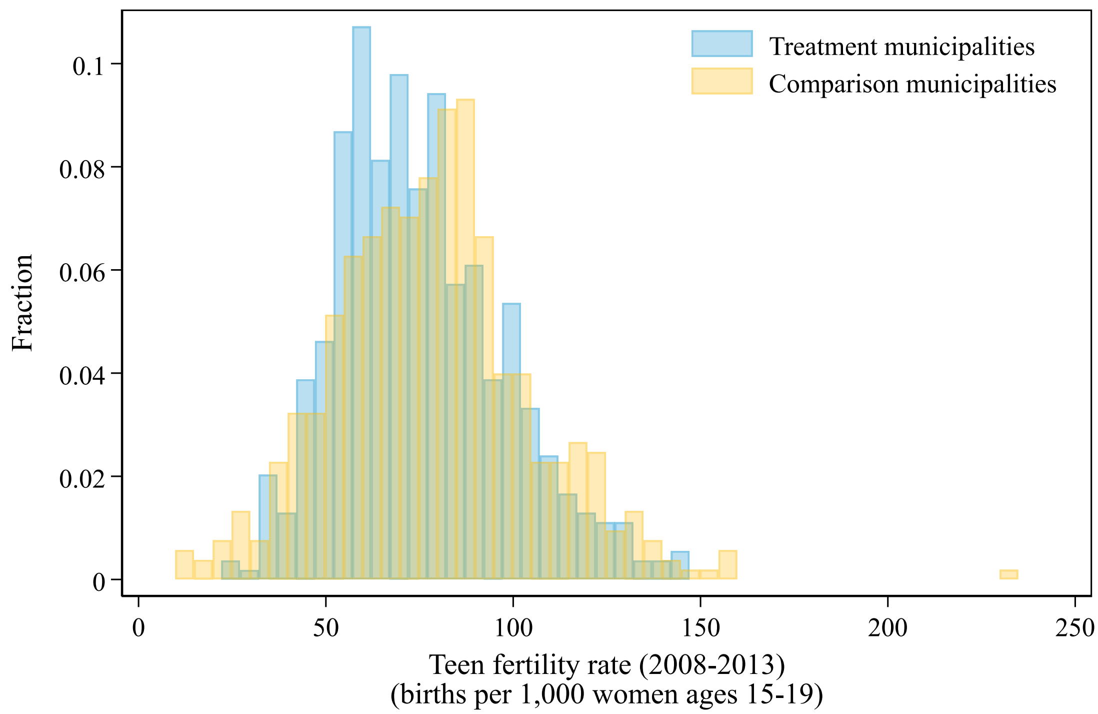
Notes: This histogram shows the distribution of SPP eligibility rates in 2014 for the 1,067 municipalities in our main sample. For the overall sample of municipalities, the minimum value is 0 and the maximum value is 285.7 (eligible female students per 1,000 female students). The median eligibility rate is 12.7. The 25th and 75th percentiles are 0 and 30.8, respectively.

Figure A4. Municipality-Level Variation in SPP Eligibility



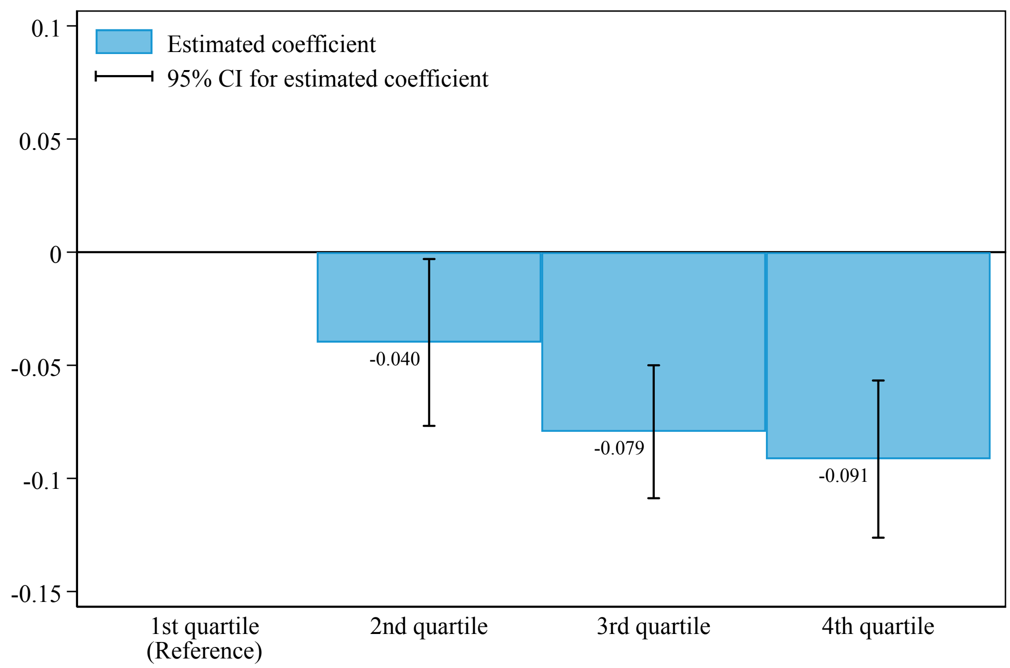
Notes: This map displays the geographic distribution of treatment and comparison municipalities. Treatment municipalities are above the median in female eligibility rates for SPP in 2014, while comparison municipalities are below the median. Dots indicate municipalities with at least one SPP-eligible HEI (red: from SPP 1; blue: added in SPP 3; green: added in SPP 4).

Figure A5. Distribution of Teen Fertility Rates



Notes: This histogram shows the distribution of teen eligibility rates in 2008-2013 for the 1,067 municipalities in our main sample separately for treatment and comparison municipalities. Treatment municipalities are above the median in female eligibility rates for SPP in 2014, while comparison municipalities are below the median.

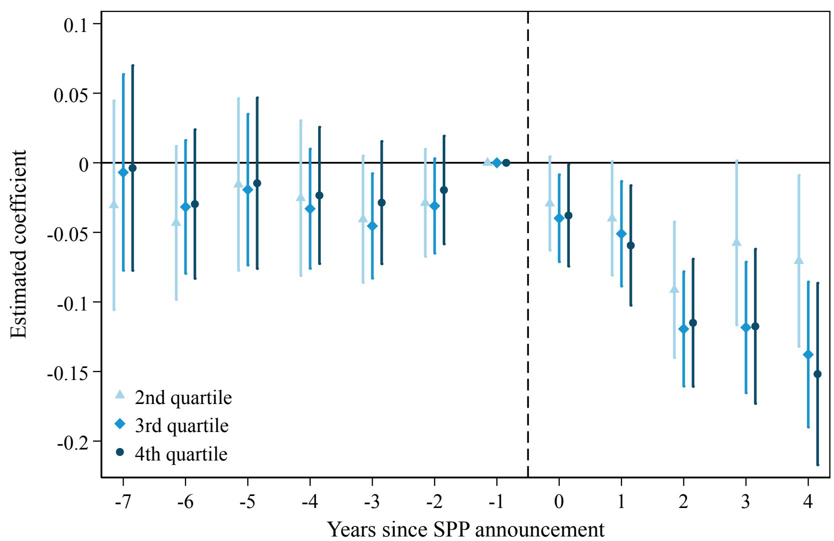
Figure A6. Summary Triple Difference Estimates by Quartile of Initial SPP Eligibility



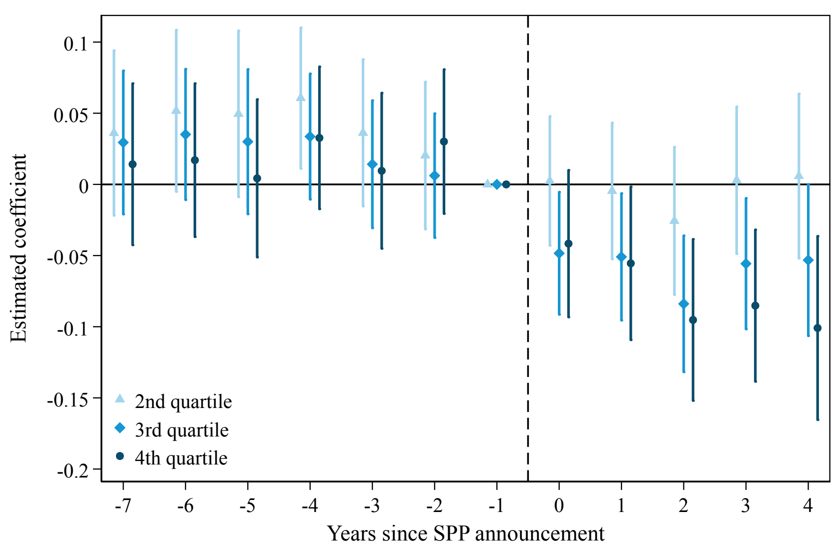
Notes: This figure plots summary triple difference estimates of β from Equation 4 using indicators for the quartile of SPP eligibility rates instead of an indicator for being above the median in SPP elgibility. All estimates are weighted by the population of women from each age group in each municipality. Standard errors are clustered at the municipality level.

Figure A7. Event Study Estimates by Quartile of Initial SPP Eligibility

(a) Difference-in-differences results

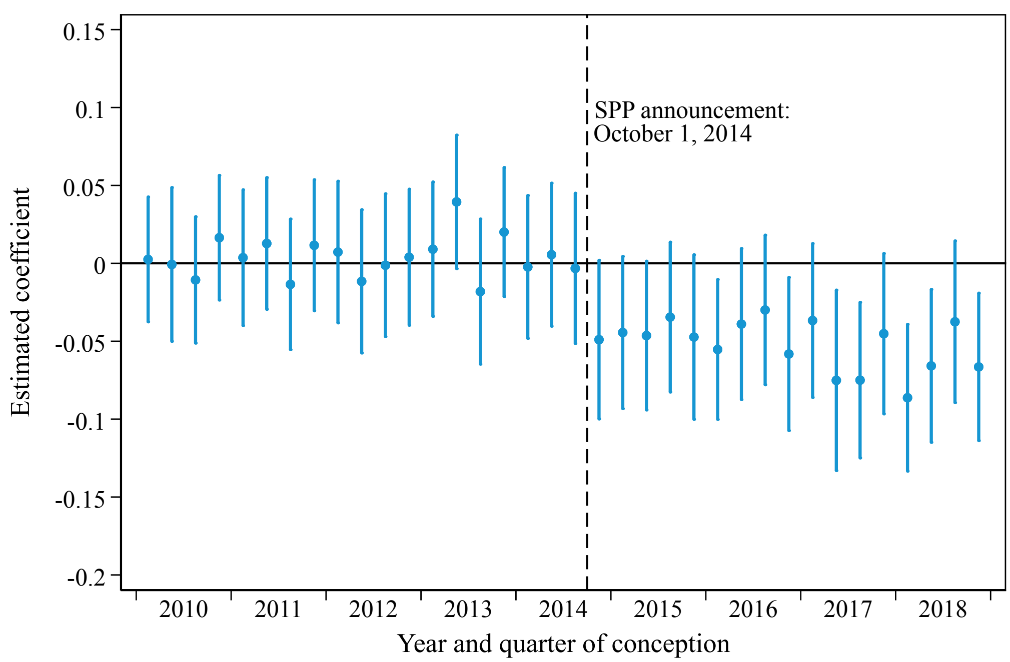


(b) Triple difference results



Notes: This figure plots the triple difference event study estimates of β_τ from Equation 2 using indicators for the quartile of SPP eligibility rates instead of an indicator for being above the median in SPP elgibility. All estimates are weighted by the population of women from each age group in each municipality. Standard errors are clustered at the municipality level.

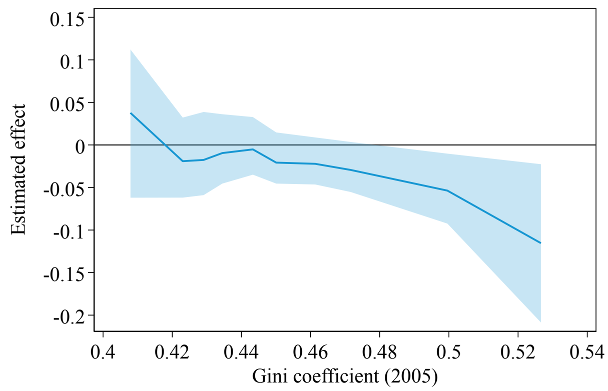
Figure A8. Triple Difference Event Study Estimates Using Quarterly Data (Poisson Model)



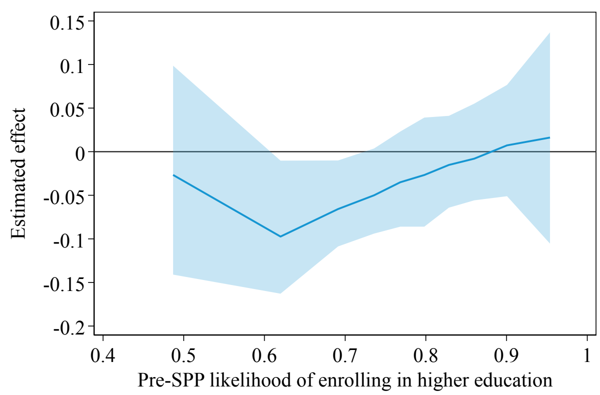
Notes: This figure plots the triple difference event study estimates of β_τ from Equation 2 using quarterly data instead of annual data. We use the birth rate as the outcome and a Poisson model instead of the log transformation with the linear model, because at the quarterly level, some municipalities have zero births in some of the cells. For estimation, we use Correia, Guimarães and Zylkin (2020)'s Stata command. The quarters in 2008 are the reference period. Only estimates for four years around 2014 are plotted. The dots represent the estimated coefficients and the vertical lines represent 95 percent confidence intervals. All estimates are weighted by the annual population of women in each municipality and age group. Standard errors are clustered at the municipality level.

Figure A9. Triple Difference Estimates by Municipality Characteristics (Non-parametric Analysis)

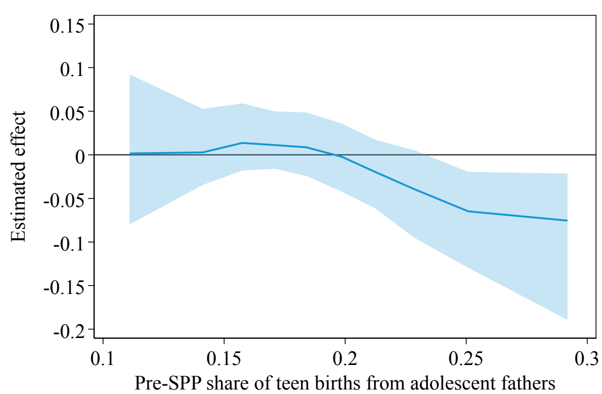
(a) Baseline Income Inequality



(b) Baseline College-Going Expectations

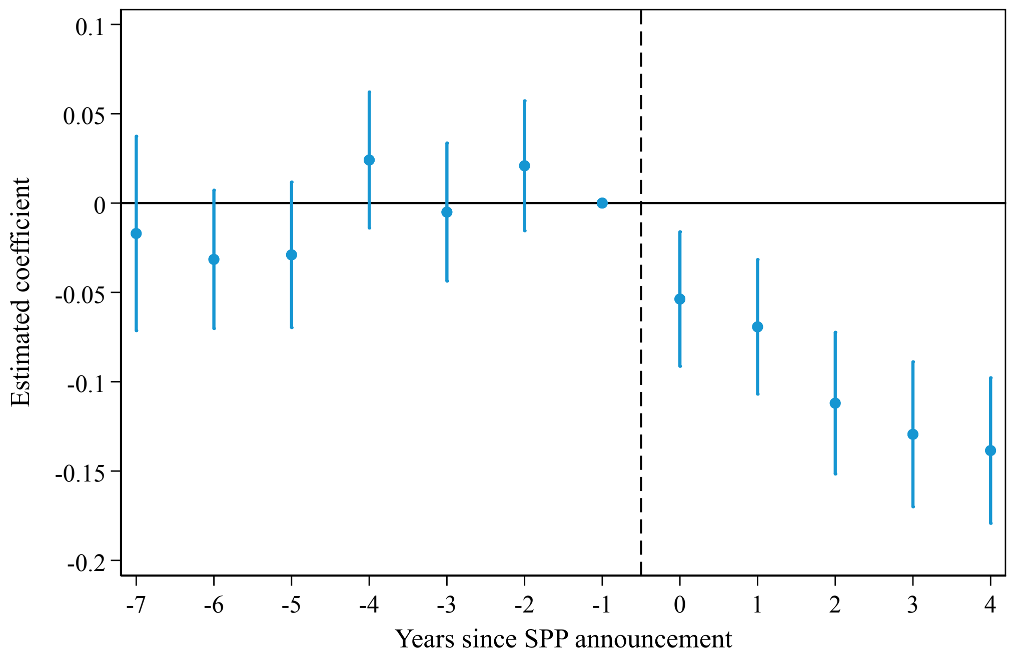


(c) Baseline Share of Teen Births from Adolescent Fathers



Notes: This figure presents triple difference estimates by municipality characteristics indicated in each panel. To do this, we first divide municipalities in deciles according to the level of each characteristic. We recast the triple difference as a simple difference-in-differences by using the within-municipality difference between the (log) fertility rate of teens and non-teens as the outcome. Then, we use Equation 3 to estimate the triple difference reduction in fertility for each decile. We then plot these estimates using a lowess regression smoothing of the triple difference coefficients on the average level of the characteristic for the deciles. The shaded area represents 95% confidence interval based on 10,000 bootstrap replications of this procedure.

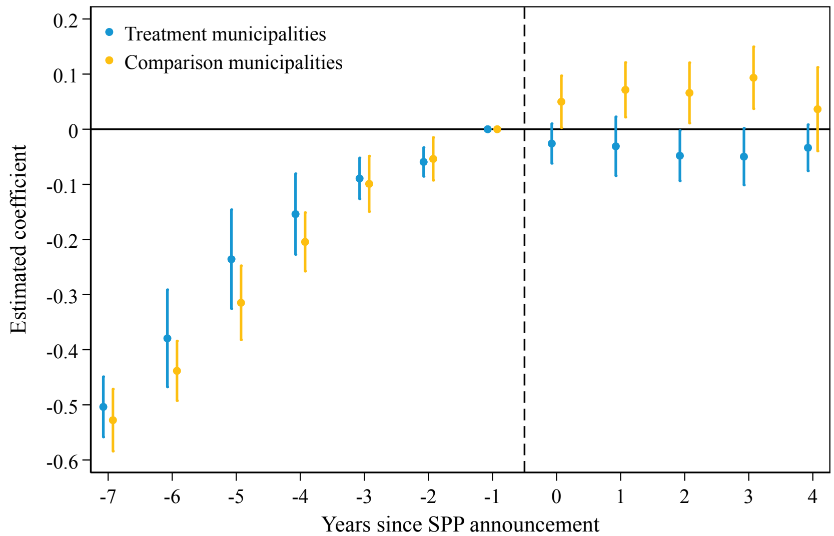
Figure A10. Event Study Estimates Exploiting Distance to Nearest SPP-eligible Institution



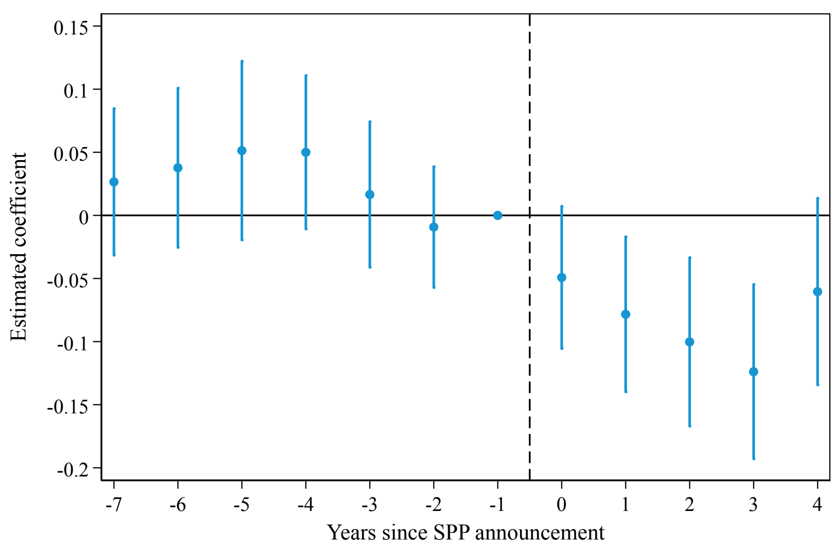
Notes: This figure plots the triple difference event study estimates from Equation 2, where SPP^*_m is replaced with an indicator for whether a municipality is below the median distance to the nearest SPP-eligible institution. The dots represent the estimated coefficients and the vertical lines represent 95 percent confidence intervals. All estimates are weighted by the number of women for each age group in each municipality. Standard errors are clustered at the municipality level.

Figure A11. Event Study Estimates Exploiting Mothers' Grade Level

(a) Trends in IHS Birth Differentials between Grade Levels



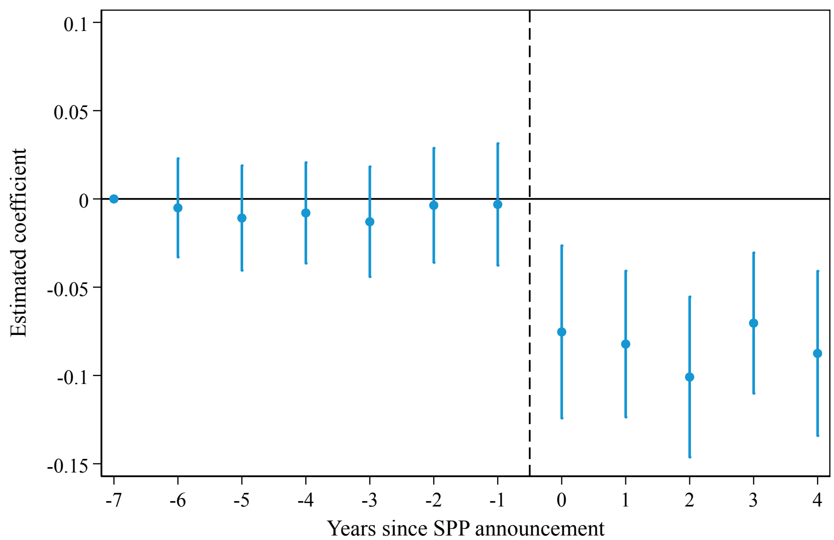
(b) Triple Difference Estimates by Grade Level and SPP Eligibility



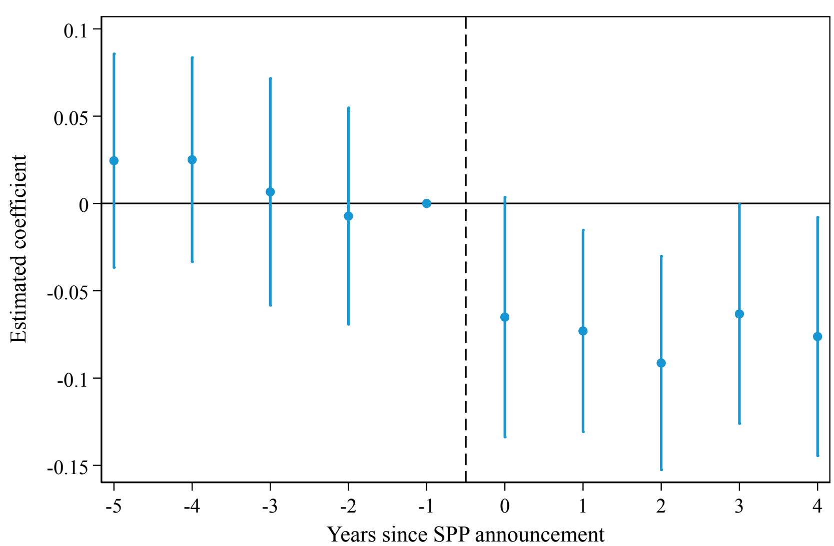
Notes: This figure plots event study estimates from Equation 2, where $Teen_a$ is replaced with an indicator for mothers who have completed eighth grade or higher. Mothers aged 15-19 with less than an eighth grade education is used as the comparison group instead of women aged 25-29. We use the number of births instead of a birth rate because we cannot reliably calculate the number of young women in each municipality with above or below an eighth grade level of education. Also, we use an inverse hyperbolic sine (IHS) transformation instead of a log transformation since some municipalities have zero recorded births in these year by grade-level cells. Panel (a) plots trends in the differentials between grade levels, separately for treatment and comparison municipalities. Panel (b) plots the triple difference coefficients. The dots represent the estimated coefficients and the vertical lines represent 95 percent confidence intervals. All estimates are weighted by the number of women for each age group in each municipality. Standard errors are clustered at the municipality level.

Figure A12. Triple Difference Event Study Estimates Using Alternative Estimators

(a) Borusyak et al. (2021)'s Imputation Estimator

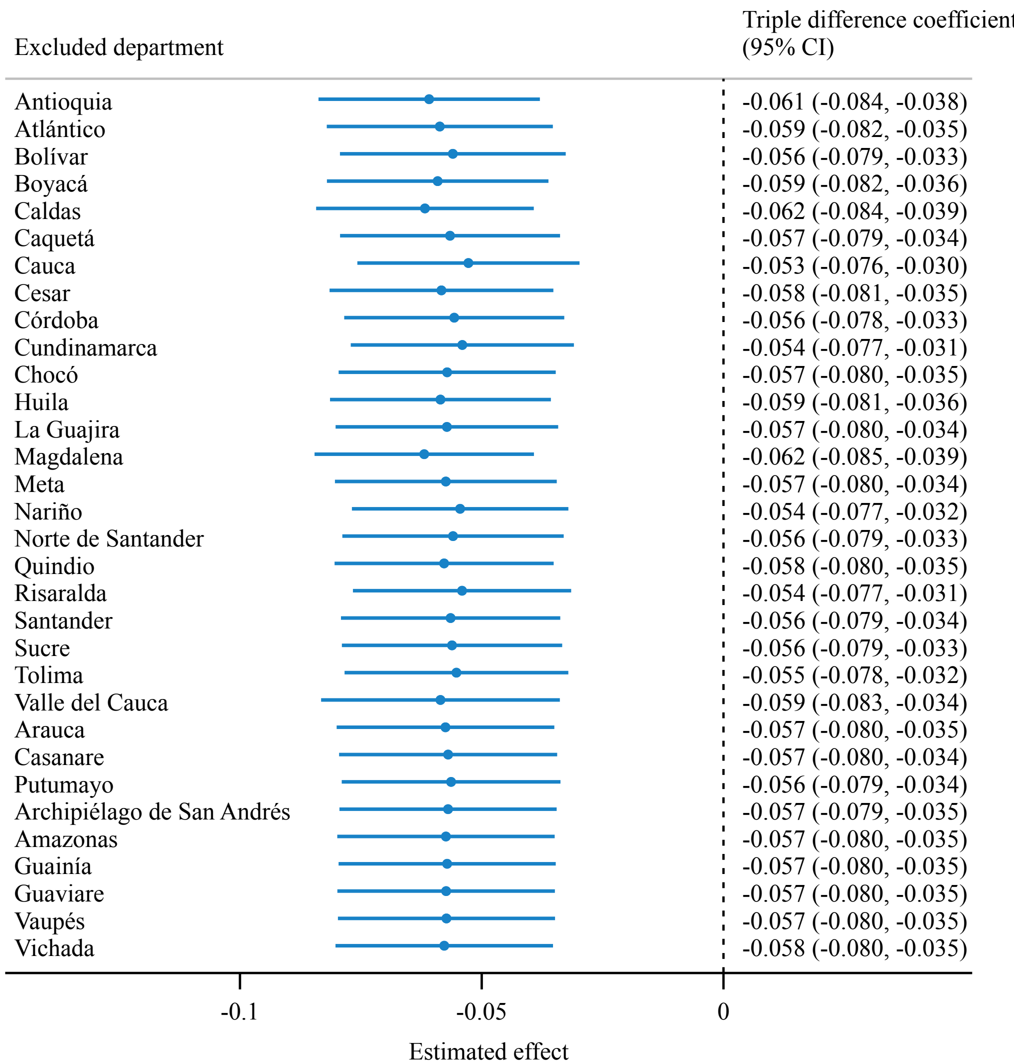


(b) de Chaisemartin and D'Haultfœuille (2022a)'s DID Estimator



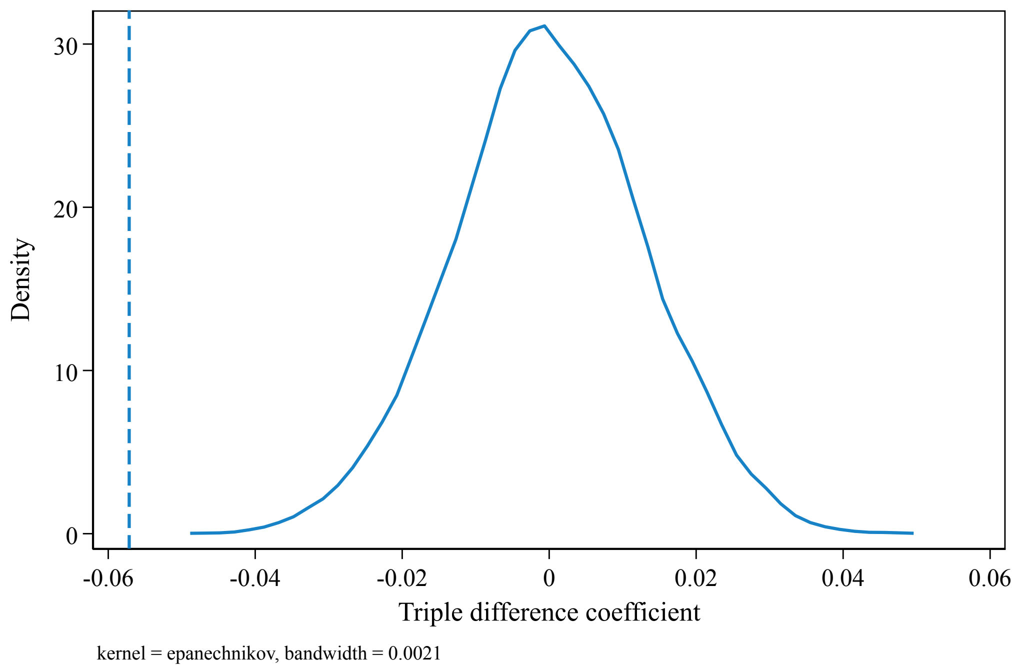
Notes: This figure plots the triple difference event study estimates using Borusyak et al. (2021)'s imputation estimator and de Chaisemartin and D'Haultfœuille (2022a)'s DID estimator. In panel (a), the estimates to the left of the dotted line correspond to OLS estimates for pre-trends and the ones to the right represent treatment effects. Vertical lines represent 95 percent confidence intervals. See Borusyak et al. (2021) for more details. In panel (b), the estimates to the left of the dotted line correspond to placebo estimates and the ones to the right represent DID treatment effect estimates. Vertical lines represent 95 percent confidence intervals. See de Chaisemartin and D'Haultfœuille (2022a) for details. All estimates are weighted by the annual population of women in each municipality and age group. Standard errors are clustered at the municipality level and are based on 5,000 bootstrap replications for the DID estimator.

Figure A13. Robustness to Excluding All Municipalities in a Given Department



Notes: This figure reports triple difference estimates of β from Equation 4 where a single department is excluded in each regression. Dots represent estimated coefficients and horizontal lines represent 95 percent confidence intervals. All estimates are weighted by the annual population of women in each municipality and age group. Standard errors are clustered at the municipality level.

Figure A14. Randomization Inference (RI): Distribution of Placebo Treatments



Notes: This figure presents the distribution of placebo treatments after 5,000 random permutations of the treatment assignment (i.e., we randomize municipalities to be treatment or comparison municipalities). We run the regressions using our summary specification in Equation 4. The vertical dashed line represents the original estimated coefficient. RI-based p-value = 0.000. The procedure was implemented using the routine by Heß (2017).

Appendix B. Validation of Treatment Intensity Measure

We attempt to probe the validity of our treatment intensity measure in the context of the previous literature by estimating whether it is associated with an increase in SABER 11 test scores after SPP is introduced. This is essentially testing whether we can replicate the results from Bernal and Penney (2019) and Laajaj et al. (2022a) using our treatment measure. We use individual-level test scores on the SABER 11 exam for 15-19 years old female students between 2010 and 2016.[37] We use a triple difference empirical approach that leverages the same municipality-level variation in SPP eligibility as in our fertility analysis (see Equation 4) and variation between students who are eligible for SPP on the SISBEN margin and those who are not.

Specifically, we estimate the following equation by OLS:

$$\begin{aligned} StdTestScore_{it} &= \phi\left(SISBEN_i^{1\text{-}2} \times SPP^*_{m(i)} \times Post_t\right) + X_{it}\Gamma_t \\ &\quad + \psi_{s(i)m(i)} + \psi_{m(i)t} + \psi_{s(i)d(i)t} + \nu_{it}, \end{aligned} \tag{B1}$$

where i denotes student, m denotes municipality, d denotes department, s denotes SISBEN level, and t denotes year. The $StdTestScore_{it}$ variable is students' SABER 11 test score standardized by test year.[38] $SISBEN_i^{1\text{-}2}$ indicates whether the student is categorized as SISBEN level 1 or 2. A SISBEN level of 1 or 2 is roughly equivalent to being eligible for SPP on the SISBEN margin, whereas students with higher SISBEN levels or not categorized are ineligible. SPP^*_m denotes treatment and comparison municipalities and is defined as $SPP^*_m = \mathbb{1}\left[SPP_m > \text{median}\left(SPP_m\right)\right]$ with SPP_m being the rate of female students eligible for the program in a given municipality.

[37] A consistent SISBEN level variable is only available in the SABER 11 data for these years. The SISBEN level is self-reported by the student. We use data from fall semesters only.

[38] The overall individual SABER 11 score is a linear combination of scores in different subjects. We follow ICFES and Londoño-Vélez et al. (2020a) (see their Online Appendix) and calculate this individual score as follows:

$$TestScore_i = \frac{Chem_i + Bio_i + Phys_i + 2SocSci_i + Philo_i + 3Lang_i + 3Math_i + Eng_i}{13} \text{ for 2010-2013, and}$$

$$TestScore_i = 5 \times \left(\frac{3Math_i + 3Reading_i + 3NatSci_i + 3SocSci_i + Eng_i}{13}\right) \text{ for 2014-2016.}$$

We then normalize these scores by year using the mean and standard deviation from the whole sample of students (males and females) in each year. We only use data from the fall semester each year.

Equation B1 contains a set of controls (X_{it}), including the average ranking for the school the student attends (as a proxy for school quality), indicators for the student's age, whether the student is enrolled in a public school, whether the student is enrolled in a rural school, the student's school schedule, and the parents' education levels. We interact all these indicators with year dummies. In the standard way for triple difference specifications, we include the three two-way interactions, denoted by ψ, between fixed effects for SISBEN levels ($s \in \{1\text{-}2, \text{Other}\}$), municipalities, and years. Similar to our main fertility specification (see Equation 4), we allow the SISBEN-specific year effects to vary by region (i.e., department). Finally, ν_{it} is an error term. In Equation B1, ϕ is our parameter of interest, measuring the effect of SPP on test scores.

Table B1. Triple Difference Estimates on SABER 11 Test Scores

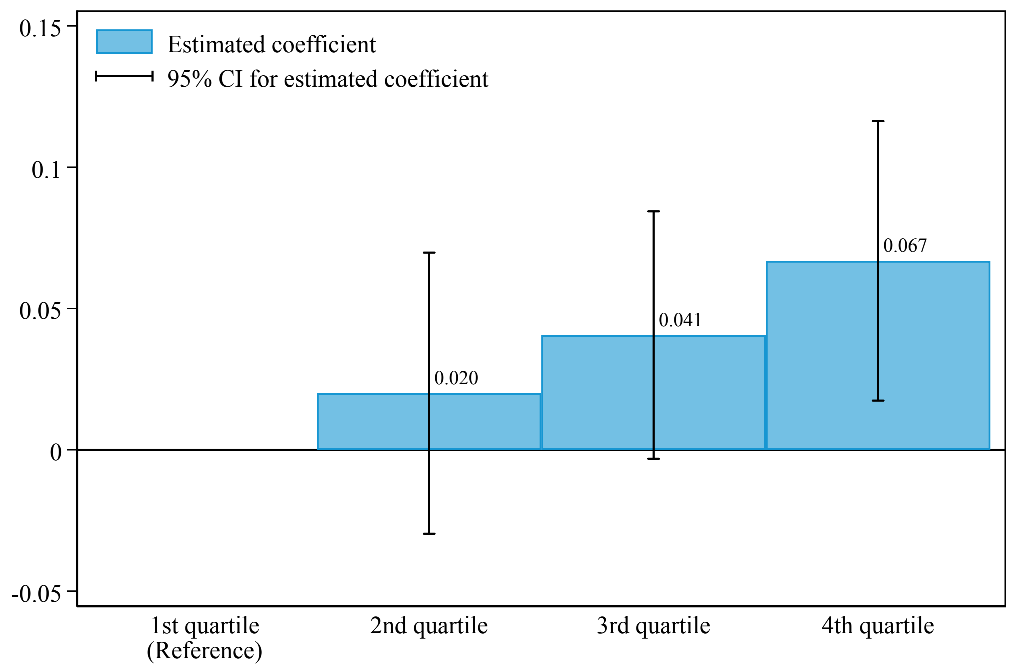
	Standardized SABER 11 test score	
	(1)	(2)
$SISBEN^{1\text{-}2} \times SPP \times Post$	0.029**	
	(0.015)	
$SISBEN^{1\text{-}2} \times SPPQuartile \times Post$		
1st quartile		[Reference]
2nd quartile		0.020
		(0.025)
3rd quartile		0.041*
		(0.022)
4th quartile		0.067***
		(0.025)
Observations	1,863,745	1,863,745
Treatment municipalities	541	–
Comparison municipalities	526	–
Pre-trends test p-value	0.055	–
Pre-SPP socioeconomic achievement gap	0.701	–

Notes: The table above reports triple difference estimates of ϕ from Equation B1. Column 2 uses indicators for the quartile of SPP eligibility rates instead of an indicator for being above the median in SPP eligibility. The reference group here are municipalities in the first (lowest) quartile of SPP eligibility. Standard errors are clustered at the municipality level and are reported in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

Table B1 presents the results from this regression. We find that, after the introduction of SPP, test scores increased for SISBEN-eligible female students in treatment municipalities by

0.029 standard deviations relative to comparison municipalities (p-value = 0.049). This increase represents about 4.1 percent of the raw pre-SPP test score gap between SISBEN levels 1-2 and higher SISBEN levels. This estimate is qualitatively similar to the estimates in Bernal and Penney (2019) and Laajaj et al. (2022a). Both Bernal and Penney (2019) and Laajaj et al. (2022a) use variations of regression discontinuity designs as their main strategies, and therefore their estimates reflect local average treatment effects. Since our triple difference estimates represent average treatment effects, it is reasonable to expect somewhat different results. However, the relative reduction in the socioeconomic gap here is remarkably similar to Laajaj et al.'s results.

Figure B1. Triple Difference Estimates on SABER 11 Test Scores by Quartile of SPP Eligibility



Notes: This figure plots summary triple difference estimates of ϕ from Equation B1 using indicators for the quartile of SPP eligibility rates instead of an indicator for being above the median in SPP elgibility. Standard errors are clustered at the municipality level.

Like for our main fertility results (see subsection 1.5), in Table B1 and Figure B1, we also report estimates where we use indicators for the quartile of SPP eligibility rates instead of being above the median SPP eligibility. A similar pattern of results emerges here. We observe bigger increases in test scores moving from the second (0.020 SD or 2.9 percent, not statistically significant) to the fourth quartile (0.067 SD or 9.6 percent, p-value = 0.008).

Appendix C. Robustness to Possibly Confounding Events

This appendix provides a full description of the analyses we conduct to assess whether events and policies that occurred around the time SPP was introduced are driving our results. We consider three events: 1) the unilateral permanent ceasefire by the Revolutionary Armed Forces of Colombia (FARC, from the Spanish acronym) in December 2014 as part of the by then ongoing peace process between the guerrilla group and the Colombian government, 2) the Zika virus epidemic, which occurred from October 2015 to July 2016, and 3) the *Jornada Única* initiative, which gradually transitioned some public secondary schools that were operating half-day shifts into full school days beginning in 2015.

Guerra-Cújar, Prem, Rodríguez-Lesmes and Vargas (2020) finds evidence that the peace agreement with FARC led to a "baby boom" in municipalities that experienced more FARC conflict before the peace agreement, and other studies find effects of the peace agreement on educational outcomes and deforestation (Prem, Vargas and Namen, 2021; Prem, Saavedra and Vargas, 2020). While Guerra-Cújar et al. (2020) find that the relative increase in fertility rates does not seem to be driven by any particular age group, we assess whether the effects of the FARC peace agreement are driving our results. We use data from Prem et al. (2020) on the locations of FARC presence in the years before the ceasefire and estimate our main specification with the subset of municipalities that did not experience any FARC-related violence in the period 2011–2014. We report these results in column (2) of Table C1. The triple difference estimate for this subset of municipalities is -0.053, nearly identical to the estimate with all municipalities.

The Zika virus can be spread from a pregnant woman to her baby, which can result in birth defects. Gamboa and Rodríguez-Lesmes (2019) studies the effect of the Zika virus epidemic in Colombia on birth rates, finding a 10 percent decline. Using municipality-level data from the Colombian National Institute of Health, we assess whether the Zika virus could be driving our results by estimating our main specification on the subset of municipalities that experienced a low incidence of Zika during 2016, the peak year of the epidemic. These results are reported in column (3) in Table C1. Our estimate for this subset of municipalities is -0.067, even larger than our main estimate. Together, these results indicate that our estimated teen fertility impacts of SPP are not driven by the FARC ceasefire or the Zika virus epidemic.

Table C1. Robustness to Possible Confounding Events

	Log fertility rate			
Municipalities:	All	No FARC	Low Zika incidence	No *Jornada Única*
	(1)	(2)	(3)	(4)
$Teen \times SPP \times Post$	-0.057***	-0.053***	-0.067***	-0.063***
	(0.011)	(0.013)	(0.024)	(0.017)
Observations	25,608	22,392	12,504	15,888
Treatment municipalities	541	485	258	304
Comparison municipalities	526	448	263	358
Pre-trends testing p-value	0.985	0.875	0.918	0.736

Notes: Column 1 reproduces the preferred triple difference results. Column 2 reports results from our summary specification in Equation 4 for municipalities that did not experience any violent events by FARC from 2011 to 2014 using data from Prem et al. (2020). Column 3 reports results from our summary specification in Equation 4 for municipalities below the median incidence of Zika in 2016. The mean incidence of Zika in these municipalities was 6.5 cases (including both confirmed and probable cases) per 100,000 inhabitants, versus 313.0 in the top half of municipalities. Column 4 excludes municipalities in which female students were exposed to *Jornada Única* ("Full School Day") at any point between 2015-2018. All estimates are weighted by the annual population of women in each municipality and age group. Standard errors are clustered at the municipality level and are reported in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

Finally, since 2015 the Colombian Ministry of Education has been gradually implementing an initiative to transition public schools from half-day shifts (morning and afternoon) to full school days to extend the duration and the quality of instruction. This policy is called *Jornada Única* or "Full School Day."[39] There is evidence from other contexts that lengthening the school day can reduce adolescent pregnancies via an incapacitation effect (Berthelon and Kruger, 2011). Accordingly, we also test that our results are robust to the expansion of *Jornada Única*. Given the high costs associated with this strategy, its expansion has been very gradual over time and was not adopted in all municipalities during our period of

[39] See Hincapie (2016) for a review of the length of the school day in Colombia around the time of the implementation of *Jornada Única*. But, shortly, in many public schools, two separate groups of students attend the same institution (i.e., use the same physical infrastructure), one in the morning and one in the afternoon. So, there are two "shifts," particularly in schools serving basic secondary (grades 6 to 9) and mid secondary (grades 10 and 11) students.

interest. In 2015, less than 0.04 percent of the female students taking the SABER 11 test attended a school with *Jornada Única*. This share increased to 0.46 percent in 2016, 6 percent in 2017, and 8 percent in 2018. We, therefore, do not expect this policy to explain the sharp decline in teen fertility observed right after the introduction of SPP in 2014. Column 4 of Table C1 corroborates this. It presents our summary triple difference estimate excluding the municipalities in which female students were exposed to *Jornada Única* at any point between 2015-2018. We still find a big, negative and significant impact of SPP on the sample of municipalities not exposed to full-day shifts due to the *Jornada Única* initiative (-6.3 percent).

Appendix D. Additional Tables and Figures for Chapter 2

Table D1. Test Scores' Role in the Gender Gap in Higher Education Enrollment (2013)

	Any HEI		High Quality HEI		High Quality HEI (Any HEI = 1)	
	(1)	(2)	(3)	(4)	(5)	(6)
Male (×100)	1.375***	-0.863***	0.938***	-0.162***	3.250***	0.115
	(0.100)	(0.185)	(0.059)	(0.054)	(0.275)	(1.099)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
SABER 11		Yes		Yes		Yes
Observations	539,997	539,997	539,997	539,997	99,913	99,913
Female mean (%)	17.8	17.8	4.9	4.9	27.5	27.5
Gap (%)	7.7	-4.8	19.1	-3.3	11.8	0.4

Notes: This table shows estimates from Equation 7. It uses data for the 2013 cohort of SABER test takers from Londoño-Vélez et al. (2020a). The outcome is an indicator variable for immediate post-secondary enrollment in a higher education institution (HEI). High-quality HEIs are those with a government-granted High Quality Accreditation. Controls include indicators for a proxy of socioeconomic status, age, mother's education, father's education, family size, attending a public school, and the school schedule. The SABER 11 control is included flexibly with indicators for performance deciles. Robust standard errors are shown in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

Table D2. Students' Characteristics

$$x_i = \beta_0 + \beta_1 NonLowInc_i + \varepsilon_i$$

x	Male		Female		Test
	β_0 (1)	β_1 (2)	β_0 (3)	β_1 (4)	(1) - (3) (5)
Age (years)	17.103	-0.386***	16.908	-0.367***	0.195***
	(0.003)	(0.006)	(0.003)	(0.005)	(0.004)
School: Public	84.753	-62.730***	87.122	-62.487***	-2.369***
	(0.083)	(0.224)	(0.070)	(0.216)	(0.108)
School: Rural	15.358	-6.877***	14.345	-7.070***	1.013***
	(0.083)	(0.163)	(0.073)	(0.143)	(0.110)
School schedule: Full day	12.780	45.861***	12.693	47.153***	0.086***
	(0.077)	(0.259)	(0.069)	(0.243)	(0.103)
School schedule: Morning	57.241	-25.684***	57.917	-27.203***	-0.676***
	(0.114)	(0.260)	(0.103)	(0.242)	(0.153)
School schedule: Afternoon	18.083	-11.742***	18.441	-11.531***	-0.358***
	(0.088)	(0.151)	(0.081)	(0.145)	(0.120)
School schedule: Evening	6.639	-4.572***	5.677	-4.313***	0.962***
	(0.057)	(0.092)	(0.048)	(0.073)	(0.075)
School schedule: Weekend	5.257	-3.863***	5.271	-4.106***	-0.014***
	(0.051)	(0.078)	(0.046)	(0.069)	(0.069)
School calendar: A	99.421	-0.410***	99.535	-0.484***	-0.114***
	(0.017)	(0.053)	(0.014)	(0.048)	(0.022)
School calendar: B	0.171	0.550***	0.125	0.655***	0.046***
	(0.009)	(0.044)	(0.007)	(0.042)	(0.012)
School calendar: Other	0.408	-0.140***	0.340	-0.172***	0.068***
	(0.015)	(0.030)	(0.012)	(0.023)	(0.019)
Mother's education: Primary or less	37.027	-27.045***	40.228	-29.667***	-3.201***
	(0.111)	(0.187)	(0.102)	(0.178)	(0.151)
Mother's education: Incomplete secondary	18.215	-10.215***	18.404	-9.589***	-0.188***
	(0.089)	(0.163)	(0.081)	(0.157)	(0.120)
Mother's education: Secondary	28.625	-2.374***	26.001	-1.515***	2.624***
	(0.104)	(0.245)	(0.091)	(0.223)	(0.138)
Mother's education: Incomplete tertiary	2.832	4.116***	2.813	4.693***	0.019***
	(0.038)	(0.133)	(0.034)	(0.130)	(0.051)
Mother's education: Tertiary	11.392	35.624***	11.054	36.255***	0.339***
	(0.073)	(0.261)	(0.065)	(0.246)	(0.098)

Notes: This table compares pre-SPP characteristics between low-income and non-low-income students, separately by gender. Each row presents results from a regression of a student characteristic (x) on an indicator for being non-low-income. Except for age, all other indicator variables denoting students' characteristics have been multiplied by 100. Columns 1 and 3 represent the averages for low-income male and low-income female students, respectively. Column 5 tests the difference between these two columns. Robust standard errors are presented in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$). Significance stars are suppressed for coefficients on the intercept term. *Continues in next page*.

Table D2. Students' Characteristics (continued)

$$x_i = \beta_0 + \beta_1 NonLowInc_i + \varepsilon_i$$

x	Male		Female		Test
	β_0 (1)	β_1 (2)	β_0 (3)	β_1 (4)	(1) - (3) (5)
Mother's education: Missing	1.909	-0.107***	1.500	-0.178***	0.409***
	(0.031)	(0.074)	(0.025)	(0.060)	(0.040)
Father's education: Primary or less	42.921	-29.781***	46.261	-32.256***	-3.340***
	(0.114)	(0.205)	(0.104)	(0.195)	(0.154)
Father's education: Incomplete secondary	16.147	-7.534***	15.140	-6.523***	1.006***
	(0.085)	(0.165)	(0.075)	(0.153)	(0.113)
Father's education: Secondary	24.214	-0.002***	22.570	1.379***	1.644***
	(0.098)	(0.237)	(0.087)	(0.220)	(0.131)
Father's education: Incomplete tertiary	2.412	3.678***	2.340	3.961***	0.072***
	(0.035)	(0.125)	(0.031)	(0.119)	(0.047)
Father's education: Tertiary	10.054	33.719***	9.336	33.864***	0.718***
	(0.069)	(0.259)	(0.061)	(0.243)	(0.092)
Father's education: Missing	4.252	-0.080***	4.353	-0.425***	-0.100***
	(0.046)	(0.111)	(0.042)	(0.101)	(0.063)
Family size: 1-2	4.703	1.749***	4.680	1.990***	0.023***
	(0.049)	(0.133)	(0.044)	(0.126)	(0.066)
Family size: 3-4	45.129	14.797***	43.532	14.683***	1.597***
	(0.114)	(0.272)	(0.103)	(0.256)	(0.154)
Family size: 5-6	37.241	-9.443***	37.503	-8.509***	-0.262*
	(0.111)	(0.251)	(0.101)	(0.238)	(0.150)
Family size: 7-8	9.298	-5.123***	10.130	-5.728***	-0.832***
	(0.067)	(0.121)	(0.063)	(0.116)	(0.092)
Family size: 9+	3.401	-1.939***	3.967	-2.427***	-0.566***
	(0.042)	(0.073)	(0.041)	(0.071)	(0.058)
Family size: Missing	0.228	-0.041*	0.188	-0.008***	0.040***
	(0.011)	(0.024)	(0.009)	(0.022)	(0.014)

Notes: This table compares pre-SPP characteristics between low-income and non-low-income students, separately by gender. Each row presents results from a regression of a student characteristic (x) on an indicator for being non-low-income. Except for age, all other indicator variables denoting students' characteristics have been multiplied by 100. Columns 1 and 3 represent the averages for low-income male and low-income female students, respectively. Column 5 tests the difference between these two columns. Robust standard errors are presented in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$). Significance stars are suppressed for coefficients on the intercept term.

Table D4. Non-linear Difference-in-Differences Estimates of the Effects of *Ser Pilo Paga* on the Gender Achievement Gap (Pooled Average)

	2015	2016	2017	2018	Average	Initial gap	Average change (%)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mean	-0.020	-0.004	-0.023	0.006	-0.012	0.134	-8.9
90% CI	[-0.041,0.002]	[-0.027,0.020]	[-0.047,0.002]	[-0.020,0.031]	[-0.031,0.007]		
95% CI	[-0.046,0.006]	[-0.032,0.024]	[-0.052,0.007]	[-0.025,0.036]	[-0.034,0.010]		
Diff-QTT(25)	0.000	-0.018	0.000	0.018	-0.009	0.054	-16.7
90% CI	[-0.038,0.038]	[-0.060,0.024]	[-0.044,0.044]	[-0.024,0.059]	[-0.043,0.025]		
95% CI	[-0.045,0.045]	[-0.068,0.032]	[-0.053,0.053]	[-0.032,0.067]	[-0.049,0.031]		
Diff-QTT(50)	-0.009	0.000	-0.018	0.036	0.009	0.134	6.7
90% CI	[-0.041,0.024]	[-0.034,0.034]	[-0.055,0.020]	[-0.002,0.073]	[-0.019,0.036]		
95% CI	[-0.048,0.030]	[-0.040,0.040]	[-0.062,0.027]	[-0.009,0.080]	[-0.024,0.042]		
Diff-QTT(75)	-0.036	-0.009	-0.062	-0.027	-0.027	0.205	-13.0
90% CI	[-0.064,-0.007]	[-0.035,0.018]	[-0.092,-0.033]	[-0.058,0.004]	[-0.051,-0.002]		
95% CI	[-0.069,-0.002]	[-0.040,0.023]	[-0.098,-0.027]	[-0.064,0.010]	[-0.056,0.002]		
Diff-QTT(90)	-0.036	0.009	-0.045	-0.027	-0.027	0.268	-10.0
90% CI	[-0.068,-0.003]	[-0.018,0.036]	[-0.073,-0.017]	[-0.056,0.002]	[-0.051,-0.002]		
95% CI	[-0.074,0.003]	[-0.023,0.041]	[-0.078,-0.011]	[-0.061,0.008]	[-0.056,0.002]		
Observations (2014)	504,171	504,171	504,171	504,171	504,171		
Observations (t)	496,949	499,562	484,347	480,600	1,961,458		

Notes: This table presents estimates from Equation 23 for selected quantiles and the mean CiC effect. The reference period ($t = 0$) for all columns is 2014. The post-SPP average estimate reported in column (5) is calculated pooling all the observations in 2015-2018 together. Confidence intervals (CI) are based on 5,000 bootstrap replications stratifying by gender, low-income status, and year.

Table D5. Falsification Test: Simulated Effects of *Ser Pilo Paga* on the Gender Achievement Gap

	2012	2013	2014
	(1)	(2)	(3)
Mean	-0.006	0.008	0.006
90% CI	[-0.023,0.010]	[-0.008,0.024]	[-0.014,0.026]
95% CI	[-0.026,0.014]	[-0.012,0.027]	[-0.018,0.030]
Diff-QTT(75)	-0.010	0.000	0.036
90% CI	[-0.035,0.015]	[-0.025,0.025]	[0.008,0.063]
95% CI	[-0.039,0.020]	[-0.029,0.029]	[0.003,0.068]
Diff-QTT(80)	-0.010	-0.010	0.018
90% CI	[-0.033,0.014]	[-0.035,0.016]	[-0.009,0.045]
95% CI	[-0.038,0.019]	[-0.040,0.021]	[-0.014,0.050]
Diff-QTT(85)	-0.010	-0.010	0.018
90% CI	[-0.035,0.016]	[-0.038,0.018]	[-0.011,0.046]
95% CI	[-0.040,0.020]	[-0.043,0.024]	[-0.016,0.052]
Diff-QTT(90)	-0.019	-0.019	0.009
90% CI	[-0.047,0.008]	[-0.051,0.012]	[-0.020,0.038]
95% CI	[-0.052,0.014]	[-0.057,0.018]	[-0.025,0.043]
Observations (2011)	485,547	485,547	485,547
Observations (t)	497,229	507,391	504,171

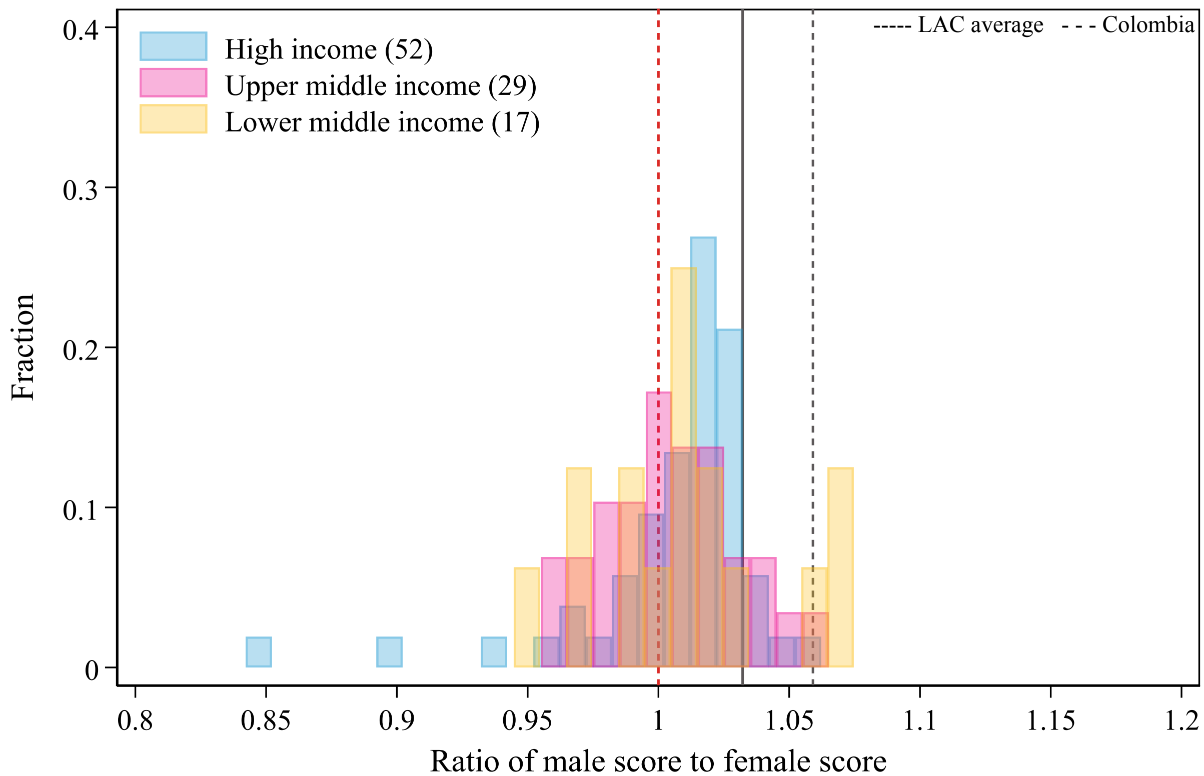
Notes: This table presents estimates from Equation 23 for years 2012-2014 using 2011 as the reference period, assuming that SPP was introduced in each of these years instead of 2014. Confidence intervals (CI) are based on 5,000 bootstrap replications stratifying by gender, low-income status, and year.

Table D6. Non-linear Difference-in-Differences Estimates of the Effects of *Ser Pilo Paga* on the Gender Achievement Gap Using Alternative Treatment Measures (2015)

	Main results	Exact SISBEN eligibility	Proxy SISBEN eligibility
	(1)	(2)	(3)
Mean	-0.020	-0.005	-0.009
90% CI	[-0.041,0.002]	[-0.019,0.008]	[-0.024,0.007]
95% CI	[-0.046,0.006]	[-0.021,0.011]	[-0.027,0.010]
Diff-QTT(75)	-0.036	-0.023	-0.009
90% CI	[-0.064,-0.007]	[-0.058,0.012]	[-0.032,0.014]
95% CI	[-0.069,-0.002]	[-0.065,0.018]	[-0.045,0.009]
Diff-QTT(80)	-0.027	-0.070	-0.027
90% CI	[-0.055,0.001]	[-0.106,-0.033]	[-0.052,-0.002]
95% CI	[-0.060,0.007]	[-0.113,-0.027]	[-0.054,0.000]
Diff-QTT(85)	-0.036	-0.023	-0.027
90% CI	[-0.066,-0.005]	[-0.058,0.011]	[-0.052,-0.002]
95% CI	[-0.072,0.000]	[-0.064,0.018]	[-0.054,0.000]
Diff-QTT(90)	-0.036	-0.047	-0.018
90% CI	[-0.068,-0.003]	[-0.077,-0.016]	[-0.044,0.008]
95% CI	[-0.074,0.003]	[-0.083,-0.010]	[-0.045,0.013]
Observations (2014)	504,171	517,565	503,150
Observations (2015)	496,949	522,558	496,067

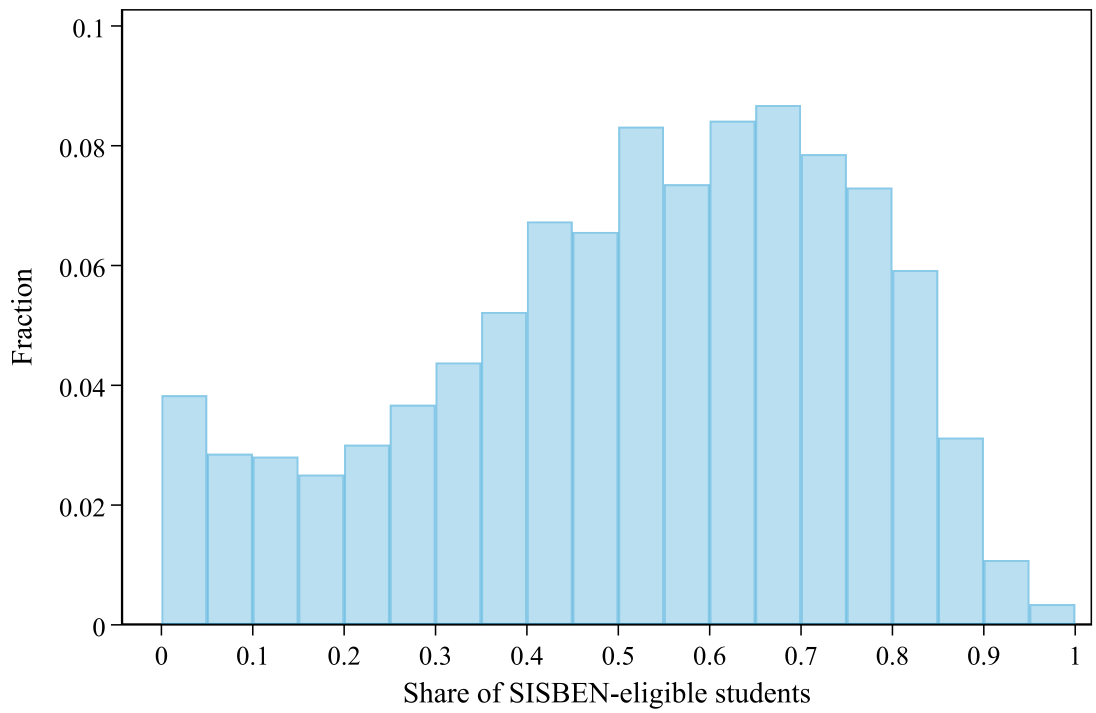
Notes: This table presents estimates from Equation 23 for the mean CiC effect and selected quantiles for 2015 using alternative definitions for *low-income* students. The reference period for all columns is 2014. Column 1 replicates the main (unadjusted) estimates. Column 2 shows estimates using data from Laajaj et al. (2022b) with actual eligibility for SPP on the SISBEN margin. Column 3 shows estimates using the primary dataset classifying as low-income to those students reporting a SISBEN level of 1 or 2, which is roughly equivalent to being eligible for SPP on the SISBEN margin. Confidence intervals (CI) are based on 5,000 bootstrap replications stratifying by gender, low-income status, and year.

Figure D1. Gender Achievement Gap in Mathematics Across Countries



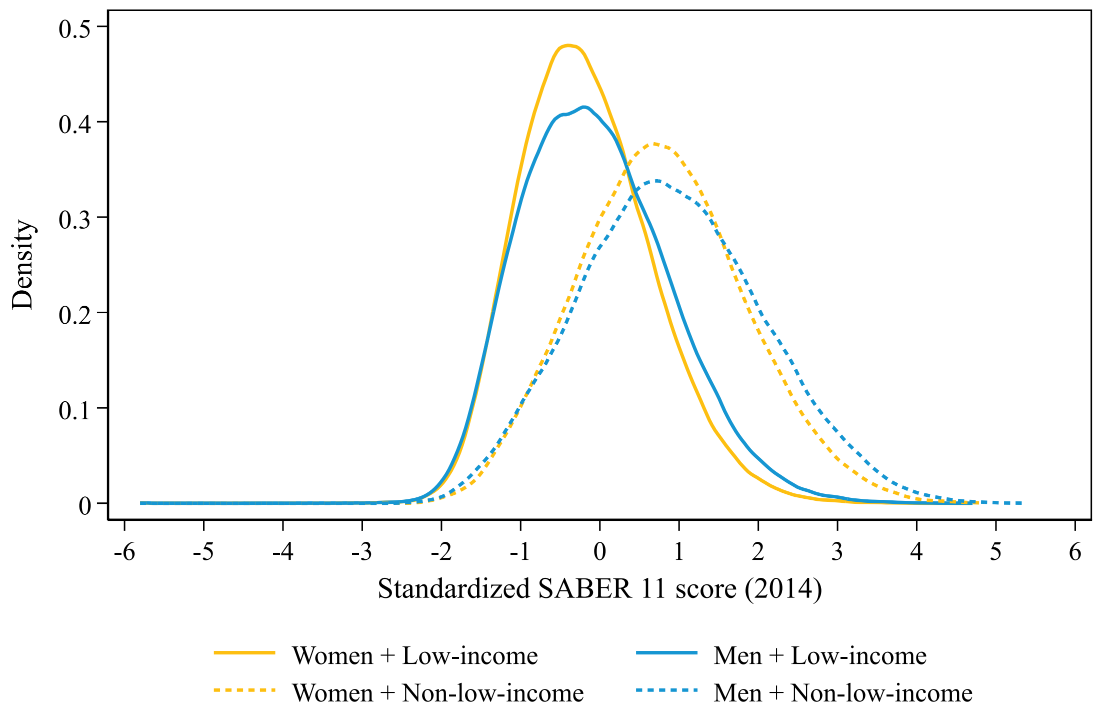
Notes: This figure presents histograms of the country-level gender gap in mathematics among secondary students in 2014 (or the latest year available before that), separately by the income group of the countries. The gender gap is measured as the ratio of the average male score to the average female score. A ratio over the unit (dashed red line) implies that men perform better than women on average. The number of countries for each histogram is indicated in parentheses. The data come from the Harmonized Learning Outcomes (HLO) database introduced by Angrist et al. (2021). The solid grey line is the simple average for countries in Latin America and the Caribbean (LAC), and the dashed grey line corresponds to the ratio for Colombia.

Figure D2. SPP Eligibility Rates on the SISBEN Margin Across Schools



Notes: This figure shows the distribution of the school-level share of students eligible for SPP on the SISBEN margin in 2014 for the 9,278 schools in the final working sample. The histogram is weighted by the number of students in each school. For the overall sample of schools, the (unweighted) median SISBEN eligibility rate is 0.587. The 20th percentile is 0.276. The figure is essencially a weighted version of Figure A3 in Laajaj et al. (2022a).

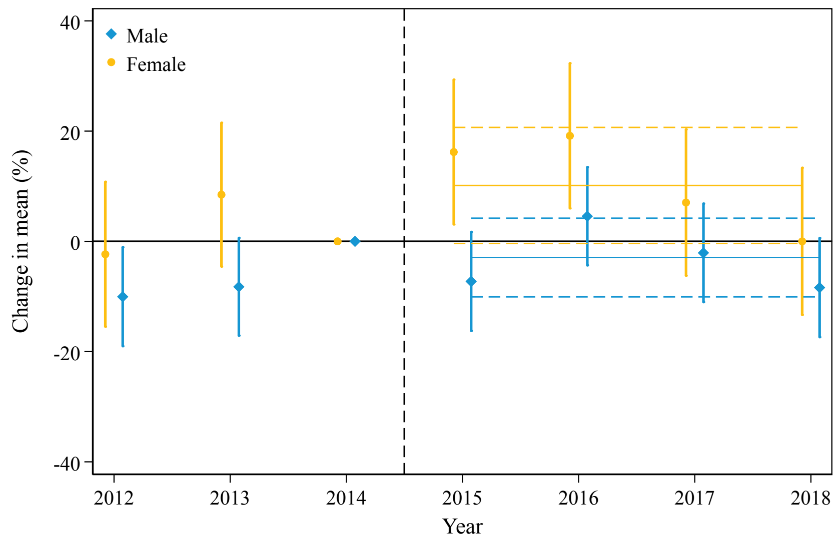
Figure D3. Distribution of SABER 11 Test Scores by Gender and Group



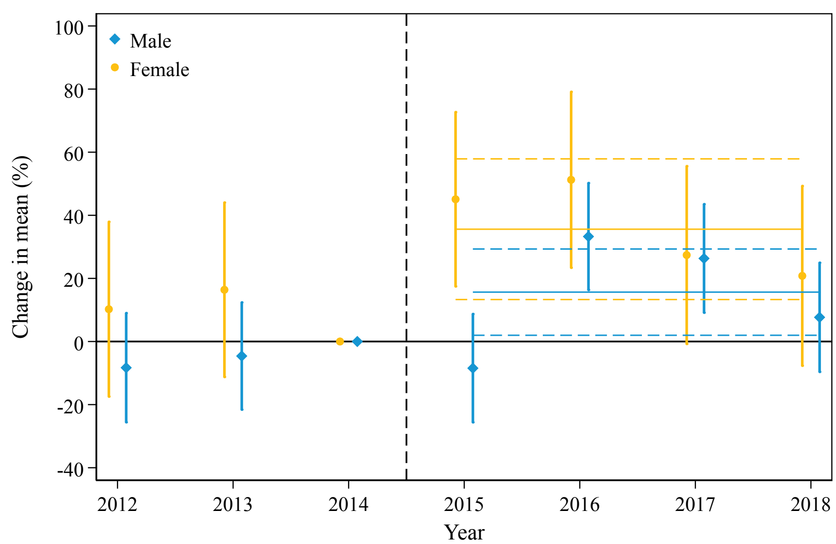
Notes: These density plots show the distribution of standardized SABER 11 test scores in 2014 for low-income and non-low-income students, as defined in subsubsection 2.4.2, separately by gender.

Figure D4. Event Study: Change in the Probability of Scoring at the Top of the SABER 11 Distribution for Low-Income Students

(a) Top 10%

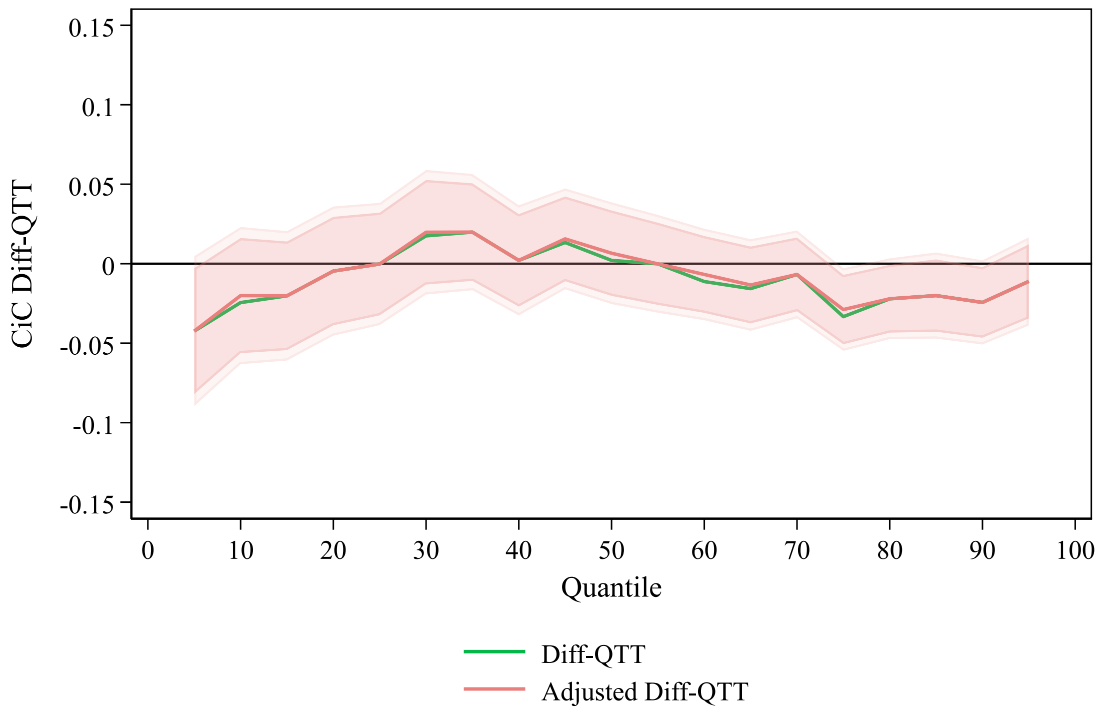


(b) Top 5%



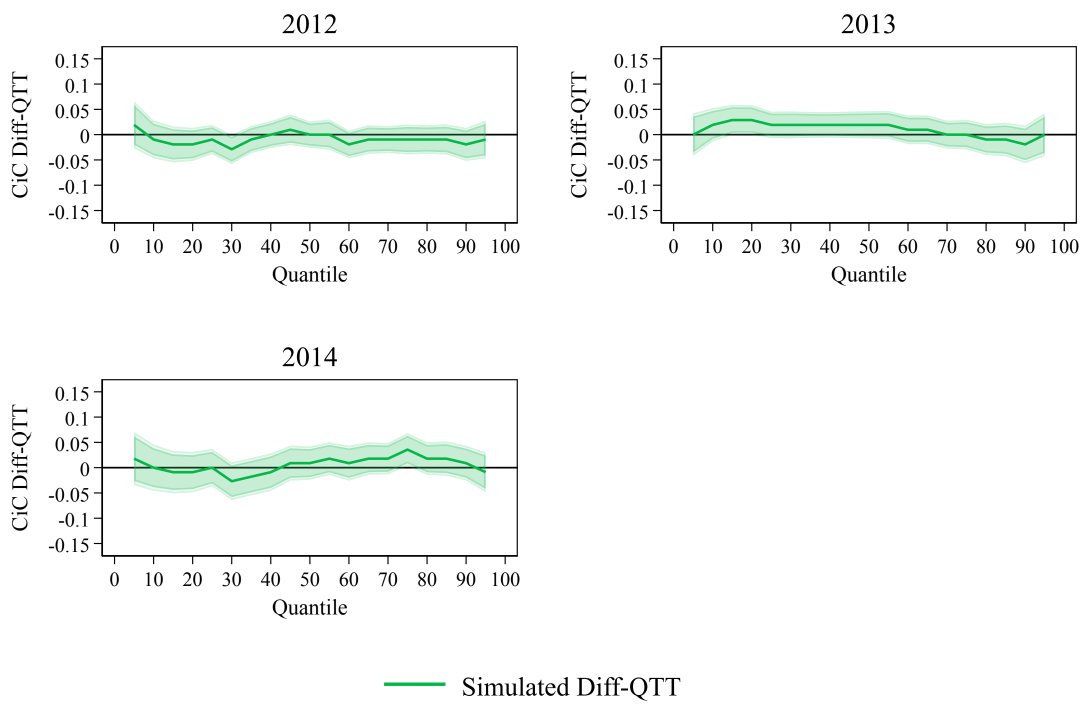
Notes: This figure shows dynamic estimates for Equation 21. It uses data from 2012 to 2018 and plots the percent change in the share of low-income students performing at the top of SABER 11 (relative to non-low-income students). The base year is 2014. Vertical lines represent 95 percent confidence intervals. The horizontal lines show the averages displayed in Table 6 and their corresponding 95 percent confidence intervals.

Figure D5. Distributional Effects of *Ser Pilo Paga* on the Gender Achievement Gap (Adjusted)



Notes: This figure presents estimates from Equation 23. For each quantile $q \in \{5, 10, \ldots, 95\}$, the solid red line reports $\widehat{\text{Diff-QTT}}^{CiC}(q)$ using the weights in Equation 26. The darkest shaded area represents 90 percent confidence intervals, while the lightest represents 95 percent confidence intervals. Confidence intervals are based on 5,000 bootstrap replications stratifying by gender, low-income status, and year. The unadjusted estimates (green line) are included for reference.

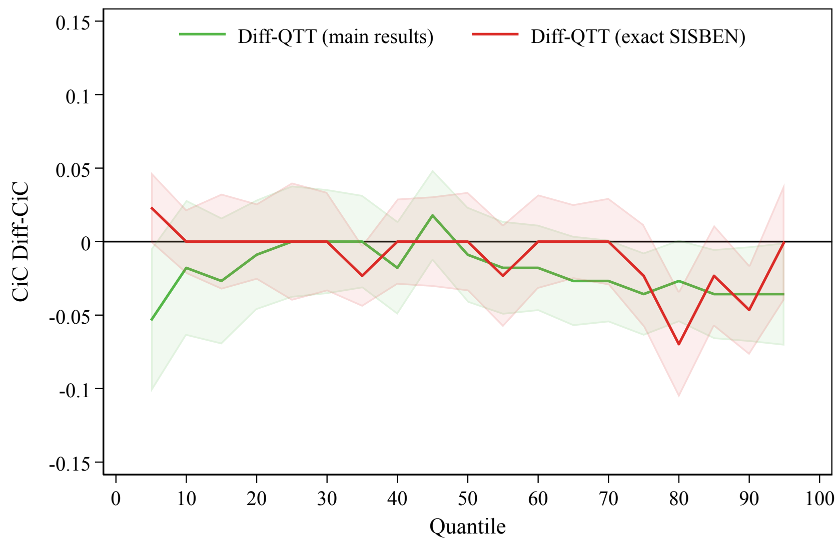
Figure D6. Falsification Test: Simulated Effects of *Ser Pilo Paga* on the Gender Achievement Gap



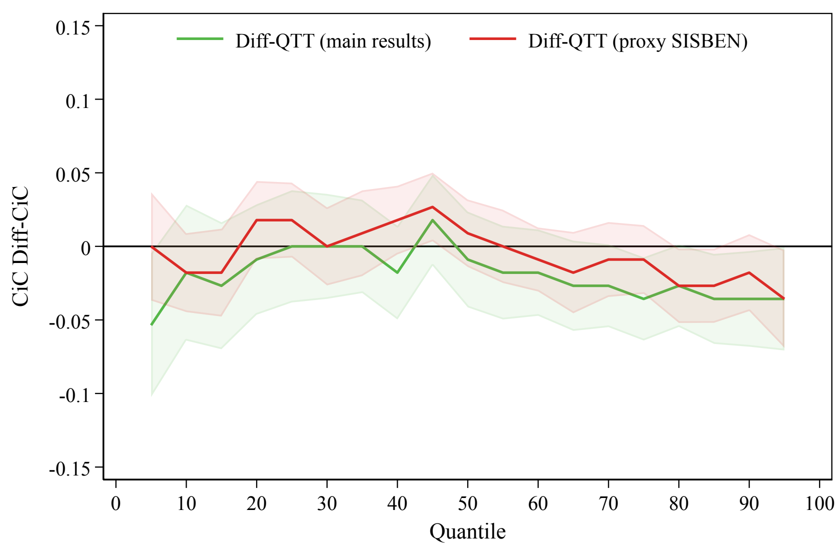
Notes: This figure presents estimates from Equation 23 for years 2012-2014 using 2011 as the reference period, assuming that SPP was introduced in each of these years instead of 2014. For each quantile $q \in \{5, 10, \ldots, 95\}$, the solid lines report $\widehat{\text{Diff-QTT}}^{CiC}(q)$. The darkest shaded area represents 90 percent confidence intervals, while the lightest represents 95 percent confidence intervals. Confidence intervals are based on 5,000 bootstrap replications stratifying by gender, low-income status, and year.

Figure D7. Distributional Effects of *Ser Pilo Paga* on the Gender Achievement Gap Using Alternative Definitions of Low-Income Students (2015)

(a) Main results vs. Exact SISBEN eligibility from Laajaj et al. (2022b)



(b) Main results vs. Proxy SISBEN Eligibility in Main Dataset



Notes: This figure shows CiC Diff-QTT estimates for 2015 from Equation 23 using alternative definitions for students. The plot above shows estimates using data from (Laajaj et al., 2022b) with actual eligibility for SPP on the SISBEN margin. The plot below shows estimates using the primary dataset classifying as low-income to those students reporting a SISBEN level of 1 or 2, which is roughly equivalent to being eligible for SPP on the SISBEN margin.

References

Abadía, Luz Karime and Gloria Bernal, "A Widening Gap? A Gender-Based Analysis of Performance on the Colombian High School Exit Examination," *Revista de Economía del Rosario*, 2017, *20* (1), 5–31.

Aizer, Anna, Paul Devereux, and Kjell Salvanes, "Grandparents, Moms, or Dads? Why Children of Teen Mothers Do Worse in Life," *Journal of Human Resources*, 2022, *57* (6), 2012–2047.

Alzúa, María Laura and Cecilia Velázquez, "The Effect of Education on Teenage Fertility: Causal Evidence for Argentina," *IZA Journal of Development and Migration*, 2017, *7* (1), 1–23.

Angrist, Noam, Simeon Djankov, Pinelopi K Goldberg, and Harry A Patrinos, "Measuring Human Capital Using Global Learning Data," *Nature*, 2021, *592* (7854), 403–408.

Athey, Susan and Guido W. Imbens, "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, 2006, *74* (2), 431–497.

Attanasio, Orazio, Lina Cardona Sosa, Carlos Medina, Costas Meghir, and Christian Manuel Posso-Suárez, "Long Term Effects of Cash Transfer Programs in Colombia," Working Paper 29056, National Bureau of Economic Research July 2021.

Avitabile, Ciro and Rafael de Hoyos, "The Heterogeneous Effect of Information on Student Performance: Evidence from a Randomized Control Trial in Mexico," *Journal of Development Economics*, 2018, *135*, 318–348.

Azevedo, Joao Pedro, Marta Favara, Sarah E. Haddock, Luis F. López-Calva, Miriam Muller, and Elizaveta Perova, *Teenage Pregnancy and Opportunities in Latin America and the Caribbean: On Teenage Fertility Decisions, Poverty and Economic Achievement*, World Bank, Washington, DC, 2012.

Badel, Alejandro and Ximena Peña, "Decomposing the Gender Wage Gap with Sample Selection Adjustment: Evidence from Colombia," *Economic Analysis Review*, December 2010, *25* (2), 169–191.

Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman, "Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa," *American Economic Journal: Applied Economics*, 2020, *12* (1), 210–259.

Becker, Gary S., "An Economic Analysis of Fertility," in Gary S. Becker, ed., *Demographic and Economic Change in Developed Countries*, Princeton, NJ: Princeton University Press, 1960, pp. 209–231.

_ , William H. J. Hubbard, and Kevin M. Murphy, "Explaining the Worldwide Boom in Higher Education of Women," *Journal of Human Capital*, 2010, *4* (3), 203–241.

Bellemare, Marc F. and Casey J. Wichman, "Elasticities and the Inverse Hyperbolic Sine Transformation," *Oxford Bulletin of Economics and Statistics*, 2020, *82* (1), 50–61.

Bernal, Gloria L. and Jeffrey Penney, "Scholarships and Student Effort: Evidence from Colombia's Ser Pilo Paga Program," *Economics of Education Review*, 2019, *72*, 121–130.

Berthelon, Matias E. and Diana I. Kruger, "Risky Behavior among Youth: Incapacitation Effects of School on Adolescent Motherhood and Crime in Chile," *Journal of Public Economics*, 2011, *95* (1), 41–53.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu, "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment," *Quarterly Journal of Economics*, 2012, *127* (3), 1205–1242.

Bharadwaj, Prashant, Giacomo De Giorgi, David Hansen, and Christopher A. Neilson, "The Gender Gap in Mathematics: Evidence from Chile," *Economic Development and Cultural Change*, 2016, *65* (1), 141–166.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes, "Staying in the Classroom and Out of the Maternity Ward? The Effect of Compulsory Schooling Laws on Teenage Births," *The Economic Journal*, 2008, *118* (530), 1025–1054.

Bloem, Michael D. and Jesús Villero, "College Opportunity and Teen Fertility: Evidence from *Ser Pilo Paga* in Colombia," Working Paper 2023.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting Event Study Designs: Robust and Efficient Estimation," 2021.

Breda, Thomas, Elyès Jouini, and Clotilde Napp, "Societal Inequalities Amplify Gender Gaps in Math," *Science*, 2018, *359* (6381), 1219–1220.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin, "Fast Poisson Estimation with High-Dimensional Fixed Effects," *The Stata Journal*, 2020, *20* (1), 95–115.

Cowan, Benjamin W., "Forward-Thinking Teens: The Effects of College Costs on Adolescent Risky Behavior," *Economics of Education Review*, 2011, *30* (5), 813–825. Special Issue on Education and Health.

Cáceres-Delpiano, Julio, Eugenio Giolito, and Sebastián Castillo, "Early Impacts of College Aid," *Economics of Education Review*, 2018, *63*, 154–166.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Difference-in-Differences Estimators of Intertemporal Treatment Effects," Working Paper 29873, National Bureau of Economic Research March 2022.

_ and _ , "Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey," *The Econometrics Journal*, 06 2022.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux, "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 1996, *64* (5), 1001–1044.

Doleac, Jennifer L. and Chloe R. Gibbs, "A Promising Alternative: How Making College Free Affects Teens' Risky Behaviors," *Unpublished manuscript*, 2016.

Duflo, Esther, Pascaline Dupas, and Michael Kremer, "Education, HIV, and Early Fertility: Experimental Evidence from Kenya," *American Economic Review*, 2015, *105* (9), 2757–2797.

_ , _ , and _ , "The Impact of Free Secondary Education: Experimental Evidence from Ghana," *NBER Working Paper*, 2021, (28937).

Dynarski, Susan, CJ Libassi, Katherine Michelmore, and Stephanie Owen, "Closing the Gap: The Effect of Reducing Complexity and Uncertainty in College Pricing on the Choices of Low-Income Students," *American Economic Review*, 2021, *111* (6), 1721–1756.

Emiliani, Laura Cepeda and Juan D. Barón, "Educational Segregation and the Gender Wage Gap for Recent College Graduates in Colombia," *IZA Discussion Papers*, 2012, (6361).

Evans, David K and Fei Yuan, "What We Learn about Girls' Education from Interventions That Do Not Focus on Girls," *The World Bank Economic Review*, 05 2021, *36* (1), 244–267.

Ferreyra, María Marta, Ciro Avitabile, Javier Botero Álvarez, Francisco Haimovich Paz, and Sergio Urzúa, "At a Crossroads—Higher Education in Latin America and the Caribbean," Report, Washington, DC: World Bank Group 2017.

Firpo, Sergio and Cristine Pinto, "Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures," *Journal of Applied Econometrics*, 2016, *31* (3), 457–486.

_ , Nicole M. Fortin, and Thomas Lemieux, "Unconditional Quantile Regressions," *Econometrica*, 2009, *77* (3), 953–973.

Flórez, Carmen Elisa and Victoria Eugenia Soto, "Tendencias del Embarazo en la Adolescencia y sus Determinantes Próximos en Colombia." In Vargas Trujillo et al., eds (2019) pp. 59–98.

Fryer, Roland G. and Steven D. Levitt, "An Empirical Analysis of the Gender Gap in Mathematics," *American Economic Journal: Applied Economics*, April 2010, *2* (2), 210–40.

Gamboa, Luis Fernando and Paul Rodríguez-Lesmes, "The Fertility-Inhibiting Effect of Mosquitoes: Socio-economic Differences in Response to the Zika Crisis in Colombia," *Economics and Human Biology*, 2019, *35*, 63–72.

García-Suaza, Andrés Felipe, Juan Carlos Guataquí, José Alberto Guerra, and Darío Maldonado, "Beyond the Mincer Equation: The Internal Rate of Return to Higher Education in Colombia," *Education Economics*, 2014, *22* (3), 328–344.

Gaviria, Alejandro, "Cambio Social en Colombia Durante la Segunda Mitad del Siglo XX," Documentos CEDE 30, Universidad de los Andes, Facultad de Economía, CEDE October 2010.

Gevrek, Z. Eylem, Deniz Gevrek, and Christian Neumeier, “Explaining the Gender Gaps in Mathematics Achievement and Attitudes: The Role of Societal Gender Equality,” *Economics of Education Review*, 2020, *76*, 101978.

Guerra-Cújar, María Elvira, Mounu Prem, Paul Rodríguez-Lesmes, and Juan F. Vargas, “The Peace Baby Boom: Evidence from Colombia’s Peace Agreement with the FARC,” *LACEA Working Paper Series*, October 2020, (0052).

Guiso, Luigi, Ferdinando Monte, Paola Sapienza, and Luigi Zingales, “Culture, Gender, and Math,” *Science*, 2008, *320* (5880), 1164–1165.

Havnes, Tarjei and Magne Mogstad, “Is Universal Child Care Leveling the Playing Field?,” *Journal of Public Economics*, 2015, *127*, 100–114. The Nordic Model.

Heß, Simon, “Randomization Inference with Stata: A Guide and Software,” *Stata Journal*, 2017, *17* (3), 630–651.

Hincapie, Diana, “Do Longer School Days Improve Student Achievement? Evidence from Colombia,” IDB Working Paper Series IDB-WP-679, Washington, DC 2016.

Husain, Muna and Daniel L. Millimet, “The mythical ’boy crisis’?,” *Economics of Education Review*, 2009, *28* (1), 38–48.

Jensen, Robert, “Do Labor Market Opportunities Affect Young Women’s Work and Family Decisions? Experimental Evidence from India,” *Quarterly Journal of Economics*, 2012, *127* (2), 753–792.

Kearney, Melissa S. and Phillip B. Levine, “Why Is the Teen Birth Rate in the United States So High and Why Does It Matter?,” *Journal of Economic Perspectives*, 2012, *26* (2), 141–166.

__ and __ , “Income Inequality and Early Nonmarital Childbearing,” *Journal of Human Resources*, 2014, *49* (1), 1–31.

Koohi, Shiva, “College Prospects and Risky Behavior among Mexican Immigrant Youth: The Effects of In-State Tuition Policies on Schooling and Childbearing,” *Economics of Education Review*, 2017, *58*, 162–174.

Kranker, Keith, “CIC: Stata module to implement the Athey and Imbens (2006) Changes-in-Changes model,” Statistical Software Components, Boston College Department of Economics June 2019.

Laajaj, Rachid, Andrés Moya, and Fabio Sánchez, “Equality of Opportunity and Human Capital Accumulation: Motivational Effect of a Nationwide Scholarship in Colombia,” *Journal of Development Economics*, 2022, *154*, 102754.

__ , __ , and __ , “Equality of Opportunity and Human Capital Accumulation: Motivational Effect of a Nationwide Scholarship in Colombia,” Replication Package, Inter-university Consortium for Political and Social Research, Ann Arbor, MI 2022. Available from `https://doi.org/10.3886/E146921V1`.

Le, Huong Thu and Ha Trong Nguyen, "The Evolution of the Gender Test Score Gap through Seventh Grade: New Insights from Australia Using Unconditional Quantile Regression and Decomposition," *IZA Journal of Labor Economics*, 2018, *7* (1), 1–42.

Londoño-Vélez, Juliana, "The Impact of Diversity on Perceptions of Income Distribution and Preferences for Redistribution," Working Paper 30386, National Bureau of Economic Research August 2022.

__ , Catherine Rodríguez, and Fabio Sánchez, "Upstream and Downstream Impacts of College Merit-Based Financial Aid for Low-Income Students: Ser Pilo Paga in Colombia," *American Economic Journal: Economic Policy*, May 2020, *12* (2), 193–227.

__ , __ , and __ , "Upstream and Downstream Impacts of College Merit-Based Financial Aid for Low-Income Students: Ser Pilo Paga in Colombia," Replication Package, American Economic Association 2020. Available from `https://www.aeaweb.org/journals/dataset?id=10.1257/pol.20180131`.

__ , __ , __ , and Luis Esteban Álvarez, "Financial Aid and Social Mobility: Evidence From Colombia's *Ser Pilo Paga*," Working Paper 2023.

McCrary, Justin and Heather Royer, "The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth," *American Economic Review*, 2011, *101*, 158–195.

Miller, Kevin, "Deeper in Debt: Women and Student Loans," *American Association of University Women*, 2017.

Molina, Oswaldo, Diego Santa María, and Gustavo Yamada, "Study for Nothing? Gender and Access to Higher Education in a Developing Country," *Economic Development and Cultural Change*, Forthcoming.

Muñoz, Juan Sebastián, "The Economics behind the Math Gender Gap: Colombian Evidence on the Role of Sample Selection," *Journal of Development Economics*, 2018, *135*, 368–391.

Muralidharan, Karthik and Nishith Prakash, "Cycling to School: Increasing Secondary School Enrollment for Girls in India," *American Economic Journal: Applied Economics*, 2017, *9* (3), 321–350.

OECD, *Education in Colombia* 2016.

__ , *PISA 2018 Results (Volume II)* 2019.

OECD and World Bank, "Reviews of National Policies for Education—Tertiary Education in Colombia," Report, Paris: OECD Publishing 2012.

Pope, Devin G. and Justin R. Sydnor, "Geographic Variation in the Gender Differences in Test Scores," *Journal of Economic Perspectives*, June 2010, *24* (2), 95–108.

Prem, Mounu, Juan F. Vargas, and Olga Namen, "The Human Capital Peace Dividend," *Journal of Human Resources*, 2021, pp. 0320–10805R2.

_ , Santiago Saavedra, and Juan F. Vargas, "End-Of-Conflict Deforestation: Evidence From Colombia's Peace Agreement," *World Development*, 2020, *129*, 104852.

Rios-Avila, Fernando, "Recentered Influence Functions (RIFs) in Stata: RIF Regression and RIF Decomposition," *The Stata Journal*, 2020, *20* (1), 51–94.

Roland, Gerard and David Y. Yang, "China's Lost Generation: Changes in Beliefs and their Intergenerational Transmission," Working Paper 23441, National Bureau of Economic Research May 2017.

Urdinola, B. Piedad and Carlos Ospino, "Long-Term Consequences of Adolescent Fertility: The Colombian Case," *Demographic Research*, 2015, *32* (55), 1487–1518.

Vargas Trujillo, Elvia, Carmen Elisa Flórez, Darwin Cortés, and Marta Carolina Ibarra, eds, *Embarazo Temprano: Evidencias de la Investigación en Colombia*, Universidad de los Andes, Ediciones Uniandes-Universidad del Rosario, 2019.

Willis, Robert J., "A New Approach to the Economic Theory of Fertility Behavior," *Journal of Political Economy*, 1973, *81* (2), S14–S64.

World Bank, *Colombia Gender Assessment*, World Bank, 2019.

Vita

Jesús Villero received his Ph.D. in economics from Georgia State University in August 2023. During his time at Georgia State, Jesús worked as a graduate research assistant with the Georgia Policy Labs. He is an applied microeconomist studying labor and development economics, with a particular interest in education and gender topics. Before joining Georgia State, Jesús worked as a research associate at Innovations for Poverty Action (IPA), where he was involved in financial inclusion projects. Jesús earned a bachelor's degree in economics from Universidad del Norte in Colombia and master's degrees in economics from Georgetown University and ILADES-Universidad Alberto Hurtado in Chile.