THREE ESSAYS ON EDUCATION DECISIONS IN COLOMBIA

BY

LEONARDO BONILLA MEJÍA

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Richard Akresh, Chair
Professor Daniel McMillen
Professor Elizabeth Powers
Professor Adam Osman

ABSTRACT

This dissertation contains three chapters that study education decisions in Colombia. Below are the individual abstracts for each chapter.

Chapter 1: Information Policies and Higher Education Choices: Experimental Evidence from Colombia

This paper studies whether providing information on funding opportunities and college premiums by degree-college pairs affects higher education decisions in a developing country. We conducted a randomized controlled trial in Bogotá, Colombia, on a representative sample of 120 urban public high schools, 60 of which were assigned to receive a 35-minute informational talk delivered by local college graduates. Using survey data linked to administrative records, we analyze student beliefs and evaluate the intervention. Findings show that most students overestimate true college premiums and are generally unaware of funding options. The talk does not affect earning beliefs but improves knowledge of financing programs, especially among the poor. There is no evidence that information disclosure affects post-secondary enrollment. However, students in treated schools who do enroll choose more selective colleges. These positive effects are mostly driven by students from better socioeconomic backgrounds. We conclude that information policies are ineffective to raise college enrollment in contexts with significant academic and financial barriers to entry, but may potentially affect certain students' choice of college.

Chapter 2: Do High School Peers Influence Post-Secondary Decisions? An Endogenous Network Approach

This paper studies the influence of high school peers on post-secondary decisions. Peer effects are identified in a social network framework. To collect information on social relationships and post-secondary decisions, over 6,000 senior-year high school students from Bogotá, Colombia, are surveyed and then followed up after graduation using administrative records. An endogenous network model is used to correct for social selection. Results indicate that close peers have some small influence on aspirations and academic performance, however, their effect is too small to translate into actual enrollment choices.

Chapter 3: Local Effects of Small-Scale Mining on School Education and Child Labor: Evidence from the Colombia's Gold Rush

Driven by a sharp rise in international prices, Colombia experienced a gold rush that reached its peak in 2012. The boom was characterized by the prevalence of small-scale artisan and illegal mining. This paper estimates the local effects of mining on schools and children. Using detailed geographic information, I construct two measures of annual change in local mining intensity capturing both legal and illegal mining: the area covered by active mining titles, and the deforestation in areas with identified gold deposits. Measurement error and potential endogeneity problems are addressed by instrumenting the mining measures with the interaction between gold prices and deposits. The main results indicate that mining significantly increases dropout rates in urban areas. For children aged 9 to 11 this is partially due to a higher probability of working. Results also indicate that in this particular context even legal mining has been harmful to children. The impact is larger when illegal mining is accounted for.

To my family.

In memory of Beatriz, who taught us so much.

ACKNOWLEDGEMENTS

The past years have been a wonderful and enriching experience, and I am most grateful to all the people who made it possible. In particular, I would like to express my gratitude and appreciations to my advisor, Richard Akresh, who always encouraged and challenged me to do better. I am also thankful to Geoffrey Hewings, Daniel McMillen, Elizabeth Powers, Adam Osman, and all the faculty members and colleagues at the Economics Department and the Regional Economics Applications Laboratory who guided me throughout the PhD program.

I would like to acknowledge and thank the Economics Department and the Colombian Central Bank for their encouragement and financial support. Special thanks to Andrés Ham and Nicolas Bottan, my co-authors in the first chapter of the dissertation, for their enthusiasm and collaboration. I am also indebted to a number of institutions and individuals, including the Colombian Institute for the Promotion of Higher Education, the Colombian Ministry of Education and the Bogota Secretary of Education, for their continuous support and access to data.

My life in Champaign-Urbana was absolutely great, and that is because I met the most wonderful people here. I am grateful to the good friends I made for understanding me and always being there with arms open wide. I will miss you very much. Thanks to the old friends as well, for their continuous support no matter the distance. And, of course, thank you Anabelle, my love and partner, for all your love and care, for your patience and good advice. This journey of ours has just begun.

Finally, I am deeply grateful to my family, to whom I dedicate this dissertation. None

of this would have been possible without your unconditional love, respect and guidance. I am blessed to have you in my life.

Contents

Chapter 1

Information Policies and Higher Education Choices: Experimental Evidence from Colombia

1.1 Introduction

Many developing countries have taken steps to reduce inequality in attendance rates for primary and secondary education. However, enrollment at post-secondary levels remains relatively low among the poor, despite its significant returns (McMahon, 2009). While credit constraints are often cited as the main barrier to attend higher education[1], recent research argues that information also plays a key role. In fact, college attendance decisions are usually based on *perceived* rather than actual net benefits (Manski, 1993a). Therefore, inaccurate beliefs may lead to sub-optimal schooling choices that have lasting consequences for lifetime earnings and welfare.

The influence of incorrect beliefs on educational choices has attracted significant atten-

[1] Previous studies suggests that liquidity constraints not only discourage potential applicants from enrolling (Manski, 1992, Solis, 2013), but also from applying for and receiving student loans (Kane, 1994, Ellwood and Kane, 2000).

tion because it has a simple and cost-effective solution: providing accurate information. At basic educational levels, the main concern is low perceived benefits of schooling. Most papers studying basic education find that students and families tend to underestimate the returns to education (Nguyen, 2008, Attanasio and Kaufmann, 2009, Jensen, 2010, Kaufmann, 2014). "Pure" information policies have proven successful in updating these beliefs. For instance, Jensen (2010) found that reading a short paragraph on the earning premiums for completing secondary increased educational attainment in the Dominican Republic by 0.20-0.35 years. Nguyen (2008) finds larger effects in Madagascar when using role models to deliver information. These treatments may achieve up to 0.24 additional years of basic schooling per US$100, which is more cost-effective than many interventions aimed at increasing schooling.[2]

Higher education schooling decisions are more complex, and so is the associated information problem. On one hand, college represents a major financial investment, and students usually have limited information regarding its costs and available funding options (Booij et al., 2012, Loyalka et al., 2013, Dinkelman and Martínez, 2014, McGuigan et al., 2014, Hoxby and Turner, 2015, Hastings et al., 2015). On the other, higher education premiums vary dramatically by college and degree, and information only recently made available to the wider public (Oreopoulos and Petronijevic, 2013, Hastings et al., 2013). While a number of countries have created websites for this purpose and encouraged stu-

[2] Cost-effectiveness calculations are taken from the Abdul Latif Jameel Poverty Action Lab website, `http://www.povertyactionlab.org/policy-lessons/education/improving-student-participation`.

dents to visit them[3], evidence suggests that they remain largely uninformed. Interestingly, many studies find that students tend to overestimate the returns to college (Pekkala-Kerr et al., 2015, McGuigan et al., 2014, Hastings et al., 2015).

This paper conducts a randomized controlled trial (RCT) in which senior high school students receive information about available funding programs and the premiums to higher education. We evaluate how providing this information affects their test scores and enrollment decisions. Our experiment takes place in public schools in Bogotá, Colombia. These schools gather students from low and middle-income families who face severe financial constraints to attend college and a very small likelihood of admission to affordable public universities. In addition, since college loans are not backed by the state, funding institutions require a co-debtor to approve any request for financial assistance. Most of the students in our sample are unable to fulfill this binding condition.

We randomly selected a citywide representative sample of 120 public schools to participate in the study. Half of these schools were assigned to receive a 35-minute informational talk delivered by local college graduates. Students were first provided with an overview of the average premiums associated to attending college compared to finishing high school (and not finishing). We then introduced the Government website where they could search for the average starting salaries of college graduates by degree-college pairs, as well as the probability of finding formal employment by degree. After this, students were briefed on

[3] Some examples are the *Observatorio Laboral* in Colombia: http://www.graduadoscolombia.edu.co, *Mi Futuro* in Chile: http://www.mifuturo.cl, and the *Observatorio Laboral* in Mexico: http://www.observatoriolaboral.gob.mx.

the admission process and availability of funding programs to cover costs. Almost six thousand students responded our baseline and follow-up surveys – the latter timed just before students sat down for the high school exit exam. Survey respondents were later matched with government administrative records that contain standardized exit exam scores and college enrollment data (degree and institution of attendance).

Our results indicate the intervention did not affect college enrollment rates. However, students in treated schools that go to college enrolled in more selective institutions. We find that these individuals increase the likelihood of enrolling in a top-10 college by almost 50% of the mean. This effect is economically significant and potentially has fairly large implications for future earnings (assuming these students graduate, of course). For instance, graduates from top-10 institutions in Colombia have a higher starting salary compared to other college graduates, about 50% on average.

The limited impact of information in increasing the demand for college may be explained by its inability to remove financial and academic barriers to entry. Most of our sample comes from low-income households, whose monthly income is unable to cover the costs of college education, has below average grades, and cannot fulfill loan requirements. In fact, students report that the most important obstacles to attend higher education are that it is unaffordable (64.5%) or difficult to gain admission (32%). Two of our results further support this interpretation. On the one hand, the information treatment increased the knowledge of funding programs but did not update earning beliefs. This is consistent with the fact that students in our sample see costs as the main barrier to attend college.

On the other hand, we find larger effects of the intervention on individuals from better socioeconomic status, for whom the likelihood of attending college is higher because these barriers are less binding.

Overall findings are consistent with existing evidence on the effectiveness of "pure" information policies for higher education. These studies provide information on costs and funding programs, college premiums, or both. Interventions focusing exclusively on costs and funding yield mixed results. For instance, Dinkelman and Martínez (2014) increase high school attendance but have no effect on academic performance in Chile. Loyalka et al. (2013) increase college enrollment despite not affecting specific college choices in China. Booij et al. (2012) find no detectable effects on loan take-up in Netherlands. Papers that only provide information about earning premiums, more in the spirit of Jensen (2010), tend to be less effective. This is the case of Pekkala-Kerr et al. (2015), who find Finnish students update their college aspirations but do not change their enrollment choices.

There are three studies similar to ours, where students receive information on premiums as well as cost and funding options. Oreopoulos and Dunn (2013) find that Canadian students raise their college earning expectations. In Avitabile and De Hoyos Navarro (2015), Mexican students improve their exit exam scores but not their dropout behavior. However, the main limitation of these two papers is that they do not assess effects on actual enrollment choices. Hastings et al. (2015) focus on a sample of students who are applying for financial aid in Chile, finding that information on costs and earnings has no effect on overall enrollment, but does encourage low-income students to choose higher-earning degrees. It is important to note that our work is different from Hastings et al. (2015) because we provide information to all students, not only those who apply for financial aid. This may be a more relevant intervention to Governments considering mass advertising of different tools to aid students in acquiring more information on college.

This study contributes to two strands of literature. First, it relates to research on unequal access to higher education. Studying how low-income students make decisions at the end of high school will shed further light on why so few apply to and ultimately enroll in college. Second, we add to the burgeoning literature that evaluates information policies at the post-secondary level, focusing on low-income students from developing countries. The findings may help understand whether an extensive low-cost information campaign is useful to attract students to college and if not, why. While our intervention is one of many possible designs, its implementation and results can potentially inform researchers and policymakers on what, how, and when information should be provided.

The remainder of this chapter is organized as follows. Section 1.2 provides background on Colombia's higher education system. Section 1.3 describes the experimental framework and intervention. Section 1.4 characterizes our data and sample. Section 1.5 presents the effects of the information treatment on higher education decisions. Section 1.6 analyzes what drives our findings by testing several mechanisms suggested by the literature, including credit constraints, gender differences, non-cognitive factors, and aspirations. We conclude in Section 1.7 by discussing our findings and outlining directions for future research.

1.2 Higher Education in Colombia

There are 327 colleges in Colombia, with 132 located in the Bogotá region.[4] Of these 132 colleges, 40 are Universities, 23 are public, and 6 are ranked top-10 in the country.[5] Degrees are classified in two levels, vocational (2-year) and academic (4-year), that encompass 55 fields. Universities supply most of the academic programs, while vocational degrees are offered at Technical/Technological Institutes. *Servicio Nacional de Aprendizaje* -SENA- is the biggest such institute in Colombia, which is public and completely free. Universities are not free, but students attending public universities pay tuition under a progressive system based on family income. While low income households pay between 0.1 and 1.8 minimum wages per semester at top-ranked public universities, the average tuition fee for private universities in the top-10 is 13.2 minimum wages.[6] High-quality public universities are highly demanded and the probabilities of acceptance are small. Scholarships for low-income students are scarce and only those who achieve the highest scores on the national exit exam have access to such opportunities.

There are two main funding programs available for the students in our sample. At the national level, there is the Colombian Public Student Loans Institution (ICETEX), an

[4] The Bogotá region includes the city and the following municipalities: Cajicá, Chía, Facatativá, Madrid, Mosquera, and Soacha.

[5] According to the 2012 Higher education exit exams (SABER PRO), the top-10 colleges in Colombia are (in order): *Universidad de los Andes*, *Universidad Nacional* (Bogotá), *Universidad del Rosario*, *Universidad Externado*, *Universidad Icesi* (Cali), *Universidad Eafit* (Medellín), *Universidad de la Sabana*, *Universidad Javeriana*, *Universidad Nacional* (Medellín), and *Universidad del Norte* (Barranquilla). *Universidad Nacional* (Bogotá and Medellín) are the only public Universities ranked top-10.

[6] Hereafter, all monetary variables will be expressed in monthly minimum wages, a commonly used measure in Colombia. The 2013 monthly minimum wage was 535,600 Colombian Pesos (roughly 288 US Dollars). The average excludes medicine, which is usually more expensive than other degrees in private universities.

agency that handles student loans for vocational, academic, and postgraduate education in Colombia and abroad. This is the largest student loan program, with 22% of enrolled students during 2013 funded by this source, and is also the most widely known. Recent reforms, that introduced zero-interest loans for low-income students, have had large impacts on enrollment and retention (Melguizo et al., 2016). The Secretary of Education of Bogotá offers a less-known funding option for low-income students from the city's public schools through the Fund for Higher Education of Bogotá (FESBO). The fund has two financing options. The first targets high achieving students and offers loans for any college or degree choice. The second only provides loans for vocational education. In both cases a fraction of the debt can be condoned if students complete the degree.

In order to obtain a loan from either funding program, students must fulfill standard application requirements. However, all credits must be backed by an approved co-debtor, a restriction that is particularly binding for low-income families. Proposed co-debtors must pass a credit check and have financial capacity to repay the full debt. In this sense, Colombia is different from Chile, which provides state-backing for college loans.[7]

There are significant differences in starting salaries for college graduates between institutions and degrees. Using official records from the Ministry of Education's Labor Observatory, which links individual-level social security records to higher education graduates, we calculate average earnings by college, degree, and field.[8] Figure 1.1 shows the distribution

[7] A more detailed description and comparison of the higher education systems of Chile and Colombia can be found in González-Velosa et al. (2015).

[8] We use the 2011 monthly salary for college graduates from 2008-2011 that report non-negative earnings.

of earnings for different categories. Notice that the choice of college matters. In fact, we observe median premiums for private and top-ranked colleges of 0.33 and 1.05 minimum wages, respectively. Degrees are at least as important. While median earnings for recent graduates with an academic degree are 2.9 minimum wages, individuals with vocational degrees make a median 1.9 minimum wages. Salaries for academic degree graduates are also much more disperse, reflecting large heterogeneity both within and between fields. This is partially confirmed by the 0.83 minimum wages premium for Science, Technology, Engineering, and Mathematics (STEM) degrees.[9]

In order to characterize the demand for higher education it is worth noting that Colombia has a large share of private high schools, particularly in urban areas. Private schools account for 28% of the class of 2013, and 51.4% in Bogotá, where higher income households opt for private education. As shown in the top-left panel of Table 1.1, 72.6% of private school students come from middle or high income families (>2 minimum wages), and 58% have at least one parent who completed higher education. In public schools, which are completely free, the share of students satisfying these two characteristics drops to 29.7% and 15.6%, respectively. One of the reasons why this happens is that private schools tend to perform better on high school exit exams and have higher college enrollment rates, particularly in selective institutions and degrees.

Test scores reflect significant differences between public and private schools. The na-

[9] Academic degrees from the following fields are classified as STEM: Agronomy, animal sciences, veterinary medicine, medicine, bacteriology, biology, physics, mathematics, chemistry, geology, business, accounting, economics, and all engineering.

tional exit exam, SABER 11, administered by the Colombian Institute for the Promotion of Higher Education -ICFES- is taken by almost every 11th-grader in public and private schools, and is required for college admission. Although the application process is completely decentralized (each institution has its own admission criteria), SABER 11 scores are heavily weighted by most universities and funding programs. Students are allowed to take the SABER 11 exam more than once, and it is relatively affordable so it is quite common to retake if necessary.[10] Over the last few years, Bogotá's private schools have consistently scored 0.76 SD above the city's public schools as Table 1.1 shows.

Less than half the students who graduate from high school enroll in college, and the odds are significantly smaller for public school students. The Ministry of Education matches the National Information System for Higher Education -SNIES- matched to the ICFES exit exam records, which allows following up students who enroll in higher education. Our estimates based on ICFES-SNIES indicate that only 46.9% of the students who graduated in Bogotá in 2013 enrolled in higher education during 2014. Moreover, private schools perform much better, since their students have consistently higher probabilities of enrolling (57.1%) and doing so in an private (42.4%) or a top-10 (16%) college. They are also much more likely to choose academic (37%) and STEM (40.8%) degrees as Table 1.2 denotes.

In summary, Bogotá has a very heterogeneous higher education system that translates into large wage premiums for selective colleges and degrees. However, there are significant financial and academic barriers to entry for low-income and low-achieving students. On the

[10] The exam fee is roughly equivalent to US$17 for students taking the SABER 11 for the first time and $21 otherwise.

demand side, Bogotá's higher income families opt for private schools that have significantly higher exit exam scores and better placement in selective colleges and degrees. This paper studies public schools in order to focus on the group that is most disadvantaged in terms of access to higher education.

1.3 Experimental Setting

1.3.1 Randomization

In order to study the effects of information on higher education decisions, we conducted a randomized control trial in Bogotá, Colombia.[11] Our population of interest were public high school students enrolled in their senior year. We focused on public schools since they have significantly lower college enrollment rates, particularly when it comes to selective institutions and degrees. A representative sample of 120 public school-shifts were randomly selected out of the 570 that offer an academic track.[12] These institutions are all mixed-sex, urban, high schools with at least 20 senior high school students enrolled in the 2012 academic year. Half of the 120 high schools were randomly assigned to receive an informational talk detailing college premiums by degree-college pairs and discussing funding opportunities, while the remaining institutions served as our comparison group.

While conducting our surveys at schools, we only interviewed students from two classrooms. These were selected at random if there were more than two classrooms at the senior

[11] This project was reviewed and approved in advance by the Institutional Review Board for the protection of human subjects of the University of Illinois at Urbana-Champaign (IRB #13570).

[12] Most public high schools in Bogotá have two shifts: morning and afternoon. Each shift has different students and most importantly, different teachers and staff. Hence, each school-shift may be considered as an independent educational institution. In what follows, we refer to school-shifts as schools.

level. Otherwise, we surveyed all students in attendance that day. In Colombia, the public school year often begins in February and ends in December. The timing of our intervention is summarized in Figure 1.2. Fieldwork for the baseline survey and the intervention took place during March 2013. The follow-up survey was conducted in August 2013, just before students took the SABER 11 exam. Our sample of schools covers a large extent of the city and most urban neighborhoods in Bogotá, with treatment and control schools being relatively spread out as Figure 1.3 shows.

1.3.2 The Intervention

During our baseline visits in March we first collected self-administered surveys. After all surveys were collected, students in treatment schools were given a 35-minute presentation delivered by young local Colombian college graduates.[13] The talk described the relationship between higher education and earnings, presented the most relevant funding programs to finance post-secondary studies, and emphasized the importance of exit exam scores for admission committees.

The talk began by describing statistics on the average monthly earnings of individuals with incomplete and complete secondary, then comparing these values to the expected salaries of individuals who completed a higher education degree (differentiating by vocational and academic).[14] We then introduced students to two websites where they could find

[13] We opted for local college graduates based on findings in Nguyen (2008), where information provided by local role models yielded higher effects.

[14] Reference earnings for incomplete and complete secondary are 0.85 and 1.07 minimum wages, respectively and were estimated using 2011 household surveys.

very detailed information on the labor market outcomes of recent higher education graduates, including average earnings by degree-college pairs and the probability of obtaining formal employment by career.[15] Additionally, we showed how the different search tools on the websites worked using some examples.

The second part of the talk focused on two funding programs: ICETEX and FESBO. For each program, we provided basic information regarding benefits, application requirements, and deadlines. Students were encouraged to visit the websites of each program for more information. We emphasized the fact that college education can be affordable, even if they choose a relatively expensive university.

The last portion of the talk focused on the importance of the high school exit exam (SABER 11). We insisted on the fact that this test is a determinant factor for admission decisions in most colleges, and that higher scores also increase the possibility of receiving funding. Students were allowed some time for questions and we gave out a one-page handout summarizing the main points of the talk and containing all the relevant links to the websites described during the talk.[16]

---

[15] The websites are: `http://www.graduadoscolombia.edu.co/` and `http://www.finanzaspersonales.com.co/calculadoras/articulo/salarios-profesion-para-graduados/45541`. They present Labor Observatory information of individuals who graduated from higher education in a user-friendly way.

[16] The original and translated copy of this handout may be found in the Appendix.

1.4 Data and Estimation Strategy

1.4.1 Data

The baseline survey collected information on 6,636 students in 116 schools.[17] The questionnaire inquired about individual demographic characteristics, family background, socioeconomic status, educational background, aspirations, current employment, future work perspectives, and attitudes towards risk. The follow-up survey was completed by 6,141 students in the same 116 schools.[18] The questionnaire followed up on some baseline questions, mainly educational and employment aspirations. It also added modules on students' household environment. In what follows, we refer to the survey data as the *Bogotá Higher Education and Labor Perspectives Survey* (BHELPS).

The survey data are further augmented by matching students in our sample to administrative sources providing information on exit exams and higher education enrollment. We match the students in BHELPS to the ICFES records, which contain scores for the high school exit exam (for the 8 different subjects and the overall score), as well as information on date of birth, gender, parents' education, and family income. We use the administrative records for these variables when they are missing in the BHELPS survey. We follow up on higher education enrollment in 2014 using the ICFES-SNIES administrative records. The matching rates for ICFES and ICFES-SNIES to the baseline sample are quite high: 95.3%

[17] Despite numerous attempts, we were unable to visit four schools. These corresponded to 3 treatment schools and 1 control school. However, the inability to interview these students does not seem to generate issues that affect randomization nor representativity as our descriptive statistics and balance tests presented below reveal.

[18] Attrition between baseline and follow-up waves was 7.5%, mainly due to absences on survey days.

and 95%. There are no significant differences between matched and unmatched students and the rates are similar across treatment and control groups.[19] We present results for three samples: i) all students observed in the follow-up BHELPS, ii) students observed in the baseline BHELPS successfully matched to the administrative data, and iii) individuals observed in the baseline and follow-up rounds of the BHELPS that are matched to the administrative data.

1.4.2 Sample Representativity and Characteristics

Our sample, which includes approximately 20% of the city's public high schools, is representative of the target population though slightly over-sampled morning-shift schools. Table 1.1 summarizes individual and school-level characteristics for all private and public schools, as well as surveyed students in the BHELPS. In Table 1.3, we present baseline characteristics for students in control and treatment groups, as well as the p-value for the differences (clustering standard errors at the school level). Both groups look very similar on their observable characteristics, suggesting that our randomization was successful.

On average, students are almost 18 years old when they graduate (measured in December 31, 2013) and most were born in Bogotá (84.7%). Almost a quarter of students have repeated at least one grade. Around 17% of students in the treatment group have at least one parent that completed college, around 2 points lower than the control group, though not statistically different. Students in treatment and control groups look almost identical in terms of family income, where 36% report income over 2 minimum wages. Since most

[19] See Table A.1 in the Appendix for attrition diagnostics.

high-income families opt for private education, we will classify students in public schools with a family income higher than 2 minimum wages as middle-income. Approximately 71% of students in the control group have internet at home, while internet access is almost 4 points lower for treatment students and the difference is barely statistically significant at the 10% level.

We asked students in the follow-up survey what they believed to be the most significant barriers to enroll in college. The majority responded that college was unaffordable (64.5%), followed by 32% who claimed that obtaining admission was the largest obstacle. This is consistent with the fact that private education is expensive and affordable public universities are very selective. While only 36% of our sample reports monthly family income above 2 minimum wages, college tuition for a semester may rise to 13.2 minimum wages at private top-10 institutions, which is equivalent to 2.2 minimum wages per month. As for progressively-priced public universities (that may cost as little as 0.1 minimum wages) admission rates are fairly low. While 40% of the students in our sample wanted to enroll in the *National University* in the baseline survey, less than 1% made it. These students might also face barriers from funding institutions. As mentioned before, most available programs require a co-debtor to back college loans.

Given that risk aversion has been found to play an important role for human capital accumulation decisions, students were asked to play two different games in the baseline.[20]

[20] Students face the following hypothetical scenario: They were just hired for a new short-term job and can choose between a fixed salary or a lottery in which earnings are determined by a coin flip. By varying the optimistic scenario payment, we classify students in a scale from 1 to 4 where 1 is extremely risk averse and 4 is risk loving. We consider a student risk averse if they are classified 1 or

The resulting classification indicated that 85% of our sample was risk averse. To measure academic self-concept, we ask students to rank themselves relative to the rest of the class on a Likert-scale from 1-10 where the latter is the highest value. As a measure of self-efficacy, students rated how often they achieved their proposed goals (from 1 to 10, where 1 is never and 10 is always). Individuals above the median response are classified as high academic self-concept and self-efficacy, while those below constitute the low group. We also asked their perceived probability of enrollment in college the following year. Almost 85% reported in the baseline survey that they were likely to enroll.

Treatment and control groups look very similar in school characteristics. Using administrative data from 2010-2012, we find on average that over 90 students per school sit for the SABER 11 exam each year. Additionally, previous cohorts performed similarly across groups. More than half the schools are morning shift and over 95% of them have a computer lab. A joint-test for balance rejects that individual and school-level attributes explain the likelihood of attending a treatment school, with a p-value of 0.680.

1.4.3 Estimation Strategy

Given the random assignment of the treatment, we quantify the effect of providing information on our main outcomes (e.g. college enrollment, SABER 11 exam scores, etc.) by estimating a cross-sectional regression, where outcomes in period $t = 1$ are explained by baseline treatment status and attributes:

2.

$$y_{is,t=1} = \alpha + \beta T_s + \theta X_{is,t=0} + u_{is,t=1} \quad (1.1)$$

where $y_{is,t=1}$ is the studied outcome for student i attending school s at the follow-up, $t = 1$. We include an intercept, α, and control for baseline student-level attributes (male, age, age squared, family income, and parental education) and school characteristics (average score on exit exam in previous years, has computer lab, shift indicators, and school size) with $X_{is,t=0}$. Our coefficient of interest is β, which captures the average effect of the informational treatment. $u_{is,t+1}$ is a mean-zero error term assumed to be uncorrelated with the treatment indicator since it was randomly assigned. Equation (3.1) is estimated by Ordinary Least Squares (OLS)[21], clustering standard errors at the school-level. Given that the actual take up of the information depends on the level of attention placed by students, β would capture the intent-to-treat rather than the average treatment effect of acquiring new information on degree-college premiums and funding options.

When studying the potential mechanisms driving our main results, we take advantage that some outcomes are available for both the baseline and follow-up BHELPS surveys. In these cases, we employ two additional specifications. First, we estimate Equation (3.1), but include the outcome at the baseline as an additional explanatory variable. This approach could potentially provide additional power. Second, we estimate a difference-in-differences specification; defining a binary variable, *Post*, that equals one after information exposure

[21] We also estimate Probit regressions but the main results are largely unchanged. We therefore choose to report only OLS estimates.

and zero otherwise:

$$y_{ist} = \alpha Post + \beta(T_s \times Post) + \mu_i + u_{ist} \tag{1.2}$$

where α estimates the change in the outcome over time and μ_i is a student-specific effect that controls for all time-invariant characteristics (observed and unobserved) in our sample. Again, β is our coefficient of interest, which measures the average effect of the information treatment on the studied outcome. Standard errors are also clustered at the school-level. Note that the modified Equation (3.1) and Equation (3.2) can only be estimated for outcomes obtained in the BHELPS surveys and not from administrative data (i.e. test scores and enrollment outcomes).

1.5 Results

This section studies the effect of information disclosure on higher education outcomes. Since information should first affect beliefs, then decisions in high school, and ultimately college enrollment, the findings are presented in that order.

1.5.1 Beliefs

Our measures of student perceptions include knowledge about funding programs and beliefs about labor market premiums. Knowledge is measured using binary variables that denote awareness of funding institutions (ICETEX and FESBO).[22] Earning beliefs are measured

[22] While desirable, we were unable to collect a measure that captures the degree of knowledge about funding programs.

by the error between perceived and actual premiums for vocational and academic degrees relative to completing high school.[23]

Baseline statistics for knowledge and beliefs are presented in Table 1.4. Almost 70% of students express familiarity with ICETEX and 18% know FESBO, with both treatment and control groups reflecting similar baseline knowledge. These patterns illustrate that students remain largely unaware of the existence of certain funding programs. On average, public high school students in Bogotá overestimate college premiums. Approximately 87.6% overestimate the premiums to vocational degrees and 89.1% for academic degrees. Reported errors for vocational and academic degrees are 69.6% and 118% larger on average. These results are consistent with findings for the same population in Colombia (Gamboa and Rodríguez, 2014) and other countries (Pekkala-Kerr et al., 2015, McGuigan et al., 2014, Hastings et al., 2015).

In addition to overestimating the average premiums to college education, students show sizable variation in their beliefs. Figure 1.4 plots the distribution of errors for vocational and academic premiums. Individuals overestimate the associated benefits of vocational degrees, but most of them are not far from the correct belief. 76.3% are within one standard deviation of the true premiums. Earning beliefs for academic degrees are more disperse: 60.1% of surveyed students have errors of one standard deviation, 29.2% between one and three standard deviations, and 10.7% more than three standard deviations. Students

[23] Similar to Hastings et al. (2015), we calculate errors by estimating the difference between perceived and actual premiums and then dividing by the actual premium. That is, if π^j denotes the wage premium and $j = \{\text{actual, perceived}\}$, then our measures are $(\pi^{perceived} - \pi^{actual})/\pi^{actual}$. Results are similar when using different measures.

are therefore more misinformed about the average premiums for academic degrees than vocational careers.[24]

Are students who overestimate different than those who underestimate? Table 1.5 presents student and school characteristics based on the direction of their baseline beliefs: below the true premium or above it. There are no differences across students in treatment and control schools, as expected. Younger students seem to overestimate college premiums for both vocational and academic degrees. Interestingly, low income students tend to underestimate the monetary benefits to college education while higher income individuals overestimate. There is also evidence that repeaters, risk averse, and more confident students are more likely to overestimate college premiums relative to their counterparts.

The effects of the information treatment on knowledge and beliefs are presented in Table 1.6. Panels A and B report cross-section estimates on two samples: all students observed in the follow-up BHELPS and students observed in both BHELPS rounds. Panel C presents the ANOVA regressions, that estimate the effect on the follow-up round controlling for the corresponding baseline beliefs. Panel D presents difference-in-differences results with individual fixed-effects. We find that the treatment increases knowledge of the largest funding program, ICETEX. Students in treated schools increase their average awareness of this institution by 3.9 percentage points, or 5.6% of the mean. The impact is larger for students observed in both rounds, with cross-section, ANOVA and difference-in-difference

[24] Jensen (2010) suggests that noisier beliefs for higher education may be due to college being a rare outcome. In our sample, less than 18% of the students have parents who completed higher education. These students have slightly more accurate beliefs for vocational degrees, but not for academic degrees compared to those whose parents have not completed higher education.

effects oscillating between 4.6 and 4.8 percentage points. Note that the sample in Panel A includes all students in the follow-up. This includes students in treatment schools who missed the presentation due to absence that day. Notice that in Panel B, when we restrict the sample to those present in the baseline (i.e. treatment students were exposed to the talk) the magnitude of the treatment effect increases. This pattern repeats itself throughout our results indicating the effect of being directly exposed to the information. Regarding the smaller funding program, FESBO, the point estimates are close to zero and not statistically significant.

We find that students are acquiring more information over time, independently from our intervention. The coefficient for the follow-up period ($Post$) in Panel C is positive and significant for both funding programs. Likewise, all individuals significantly reduce the degree to which they were overestimating college premiums. This reflects that students in our sample gain further knowledge about higher education during their senior year.

One potential reason we do not find that students in treated schools corrected their beliefs at a faster rate than control students could be due to opposing effects: students who were initially overestimating before the intervention update downwards and those that were underestimating update upwards. We test for this possibility by estimating separate regressions for each group defined at baseline in Table 1.7. Similar to the average effects, individuals do correct their beliefs in the appropriate direction, but not because of the information treatment. Once again, students acquire information over time on their own, pushing them closer to the actual earning premiums.

As an additional robustness test, we change the reference values for earning beliefs. In all previous estimates, we compared students' perceptions to the average vocational and academic premiums with respect to high school. Perhaps students used their own expectations as a reference instead of those for an average individual. In the baseline BHELPS, we asked students to tell us the degree, college, and field they aspired. Using the records from the Labor Observatory on starting salaries for college graduates, we calculated two measures of expected earnings for each student: i) by degree and field, and ii) by degree and college. The same analysis from Tables 1.6 and 1.7 confirms that the treatment did not affect premium beliefs (results are shown in Table A.2 in the Appendix).

1.5.2 Test scores

As previously mentioned, academic performance plays a central role in college admissions in Colombia. The informational talk could have affected effort in high school by increasing the desirability or attainability of a post-secondary degree. We measure student performance using test scores from the national high school exit exams (SABER 11) that was taken approximately five months after our intervention. In particular, we focus on the overall score and the two most important subjects: mathematics and language.[25] All scores are standardized with mean zero and standard deviation of one with respect to the control group for ease of comparison.

Table 1.8 presents the average effects of information on test scores for all students

[25] The overall score is computed using the official weights: mathematics (3), language (3), social sciences (2), biology (1), physics (1), chemistry (1) and philosophy (1).

matched to administrative records (Panel A), and two more restricted samples of students: those observed in the baseline BHELPS that were successfully matched and individuals observed in both baseline and follow-up who were matched (Panels B and C). While the estimated coefficients are consistently positive for mathematics, we do not find statistically significant effects of the treatment on test scores for any sample, on average. We also test for differential effects along the score distribution using quantile regressions finding similar results (see Figure A.1 in the Appendix).

1.5.3 College Enrollment

We are able to track students who enrolled in higher education after graduation, and may further characterize their college and degree of choice. The enrollment rate for a post-secondary degree (academic or vocational) in our sample is 44%, with around 34.6% enrolled in a vocational program. Less than 10% of the students enroll in academic degrees, very few in top-ranked colleges (1%), and STEM degrees (4.9%).

Table 1.9 presents treatment effect estimates on higher education enrollment for the same three samples used in Table 1.8. We find that the effect of information on the probability of enrolling in any post-secondary program is positive, though not statistically distinguishable from zero. We do find a positive and statistically significant effect on the probability of enrolling in a top-10 college. These effects range from 0.4 to 0.6 percentage points depending on the sample. This impact, though small in magnitude, is also economically significant. In fact, it represents an increase of approximately 50% with respect to the control group's average. Estimated effects on the other three intensive margin outcomes are also positive but not statistically significant.

Our results are consistent with previous literature. Among "pure" information treatments, most studies find no effect of disclosing information on higher education enrollment (Booij et al., 2012, Fryer, 2013, Oreopoulos and Dunn, 2013, Pekkala-Kerr et al., 2015, McGuigan et al., 2014, Dinkelman and Martínez, 2014, Wiswall and Zafar, 2015). Our intensive margin effects are similar to those of interventions focusing on students who are already applying to college and have a high probability of enrollment (Hoxby and Turner, 2013, Hastings et al., 2015). In the long run, opting for a top-10 college may have important implications on future earnings (conditional on graduating). Recall from Figure 1.1 that students who graduate from a top-10 college in Colombia earn approximately 50% more than non-top college students (1 minimum wage more on average). Therefore, while providing information may not lead more individuals to attend college, it does seem to affect what colleges are chosen by those who do enroll.

1.6 Mechanisms

The effects of providing "pure" information appear to have been modest overall. On the one hand, students update their knowledge on funding programs but not their earning beliefs. On the other hand, we observe no improvement on college enrollment but a higher likelihood of attending top-10 colleges. In this section we explore potential mechanisms that help interpret these results.

Our analysis highlights the role of credit constraints and gender differences. We have already discussed that the main barrier to college attendance for low income students in Colombia are its high costs. Additionally, there remain considerable gender differences in higher education choices and labor market outcomes (Goldin et al., 2006). In part, this may reflect gender-specific traits or preferences that affect boys and girls differentially.[26] Complementary to this mechanism, we explore the role of non-cognitive factors and post-secondary aspirations.[27] In the following section we estimate heterogeneous effects to examine the degree to which the information treatment could have affected groups in different ways. For example, the presentation could have discouraged poorer students while encouraging wealthier students to attend college, and the resulting average effect could be zero in such an event.

1.6.1 Credit Constraints

To evaluate the extent to which credit constraints could explain our results, we explore the heterogeneity of treatment effects by estimating fully interacted versions of Equation (3.1) by income groups. Table 1.10 presents the treatment effects for each group (low and middle income) for our main outcomes. It also includes the p-value for a Wald test that

[26] For instance, there is evidence that when given the option, women shy away from competition (Niederle and Vesterlund, 2007), perform less well in competitive environments (Gneezy et al., 2003), and self-select into less competitive or lower earning careers (Buser et al., 2014).

[27] We focus on three non-cognitive dimensions that have been identified as critical determinants of human capital accumulation and academic success: risk aversion (Belzil and Hansen, 2004, Belzil and Leonardi, 2007, Heckman, 2007), self-concept and self-efficacy (Bénabou and Tirole, 2002, Heckman et al., 2006). Aspirations may keep students from pursuing more ambitious goals or induce frustration because of the difficulties in achieving their them (Appadurai, 2004, Ray, 2006, Heifetz and Minelli, 2014, Genicot and Ray, 2014, Dalton et al., 2016).

these coefficients are equal. For parsimony, we focus on the sample of students observed in the baseline that are matched to the later rounds of survey data and administrative records.[28]

In column (1) we find that only students from low-income families learn about ICETEX – the main funding institute. While the estimated effect on middle-income students is not statistically significant, low-income students increased their knowledge of ICETEX by about 6.1 percentage points. This effect is statistically different and more than twice than the effect on middle-income students. This likely reflects a catching-up: students from higher income families report significantly higher knowledge of funding programs in the baseline survey. We do not find any statistically relevant effects or differences between income levels for all other knowledge or belief outcomes. Overall, students appear to have valued information on financing more than that of earnings, suggesting that credit constraints are indeed a primary concern for most of the students in our sample.

Columns (2) to (4) present heterogeneous effects of the intervention by income level on test scores. We find that students from middle-income families increase their test scores significantly more than those from low-income families. The estimated coefficients on math for middle income students oscillate between 7.6% (p-value=.100) and 9.3% standard deviations and are statistically significant for individuals observed in both rounds of the BHELPS survey (Table A.3).

[28] Appendix Table A.3 presents results using individuals observed in both rounds of the BHELPS and matched to each source of administrative data. Those findings are unchanged from those discussed here.

Heterogeneous effects by family income on enrollment outcomes are presented in columns (5) to (9). As with the average estimates, we find that all coefficients are positive though generally not statistically significant. However, students' intensive margin decisions respond differently to information depending on their income category. First, the effects previously found on entry to top-10 colleges is driven by middle-income students. The estimated effect is 1.7 percentage points and statistically different from that of low-income students. Second, poorer students increase their probability of enrolling in a private college, with an estimated coefficient of 2.1 percentage points (though the difference with respect to middle-income students is not statistically significant).

In general, we find that most of the positive effects of the intervention were on the students from middle-income families in our sample. This further supports the idea that providing information may have limited effects on higher education demand when such interventions do not eliminate the main barriers to entry. In the Colombian case these are twofold: sizable credit constraints and low probabilities of admission to affordable institutions. Since most of the higher-income students are already aware of available funding options, it seems plausible that information provides them with additional motivation to perform better on the exit exam and therefore attend more selective colleges.[29]

[29] In the Appendix, we also consider heterogeneous effects by the direction of errors in baseline earning expectations. Our results showed that poorer students underestimate college premiums while richer children overestimate. Findings are shown in Table A.4 and are similar to those using income groups. While information has slightly larger positive effects on those who underestimate, these differences are not statistically significant.

1.6.2 Gender differences and Other Factors

In Panel B of Table 1.10, we present heterogeneous effects by gender. At baseline, boys had significantly lower knowledge of ICETEX than girls. The treatment appears to have bridged this gap as suggested by a positive effect on males of 6.9 percentage points, which is statistically distinguishable from females. At the same time, though estimated effects on test scores are positive throughout for boys and mostly negative for girls, these effects are not statistically different to zero.

Evaluating the heterogeneous effects on enrollment outcomes in columns (5) to (9), we find suggestive evidence that the information treatment increased private, and top-10 college enrollment for boys, but no statistically significant effects for girls. Nevertheless, we cannot reject the null hypothesis that the effects between males and females are the same. Overall, it seems that other than catching up on knowledge of the main funding programs, the treatment did not significantly affect boys and girls differently. If anything, it may have encouraged boys slightly more than girls to pursue a degree in a more selective institution.

Other than gender, non-cognitive factors may also play an important role in determining human capital accumulation and academic success. In Table A.5 in the appendix, we assess potential heterogeneity by three non-cognitive dimensions: risk aversion, self-concept and self-efficacy. We also estimate differencial effects by perceived likelihood of enrollment, which reflects not only students' self-concept and self-efficacy, but also accounts for the financial constraints they foresee. We find that treatment effects on test scores and college and degree choice are concentrate on students with high-efficacy, low risk-aversion and high perceived likelihood of enrollment. Finally, we examine whether information affects student aspirations.[30] As can be seen in Table A.7 of the Appendix, there are positive and significant effects on academic and STEM degrees in the cross-section and ANOVA estimates, but they are not statistically significant in the difference in difference estimation. This suggests that intensive margin effects on enrollment are not driven by changes in student aspirations.

1.7 Conclusion

This paper analyzes whether providing information on funding opportunities and college premiums by degree-college pairs affects higher education decisions in Bogotá, Colombia. We conduct a randomized controlled trial on a representative sample of 120 urban public high schools, half of which received an informational talk. Using survey data linked to administrative records, we analyze student beliefs and evaluate the intervention. We find that most students overestimate true college premiums and are generally unaware of funding options. The talk does not affect earning beliefs but improves knowledge of financing programs, especially among the poor. There is no evidence that our treatment affects post-secondary enrollment. However, students in treated schools who do enroll choose more selective colleges. These positive effects are mostly driven by students from better

[30] We exploit a question in both BHELPS waves that asks students what college and degree they would like to attend. Descriptive and balance statistics for the aspiration outcomes may be found in Appendix Table A.6.

socioeconomic backgrounds.

Our findings confirm that misinformation is a problem among potential college entrants since they tend to overestimate its benefits and are mostly unaware of its costs. However, this is not the main deterrent for attending college. The existence of significant academic and financial barriers to college entry in Colombia might limit the influence of better information because low-income students believe the system limits upward mobility. In fact, we find larger effects of the intervention on middle-income individuals, for whom the likelihood of attending college is higher since constraints are less binding. Moreover, our treatment increased the knowledge of funding programs but did not update earning beliefs. This is consistent with most students in our sample believing that costs are the main barriers to higher education. We conclude that providing information cannot single-handedly increase higher education enrollment among low-income students in this context. It takes more comprehensive measures, such as zero-interest rates loans (Melguizo et al., 2016), to achieve substantial improvements in this respect.

Despite the inability to attract more low-income students into college, providing information has some positive effects on college choices for those who enrolled. These results are particularly interesting since we targeted a wider population than other papers, such as Hastings et al. (2015) and Hoxby and Turner (2013), and yet found similar results in the intensive margin. Given the low-cost of "pure" information interventions, policymakers may therefore consider less targeted policies to orient students in their college choices, even if only a fraction of them is expected to benefit from the additional information.

How and when to provide information is an interesting direction for future research. Our intervention is one of many possible designs in this respect. For instance, while we provided average college premiums, future studies could present the entire distribution of earnings in a simple and intuitive manner. Likewise, disclosing more detailed cost data may be useful. The timing of information policies, especially for higher education choices, is also highly relevant. Additionally, whether these interventions should target students, parents, or both is an open-ended question. Our results indicate that providing information to students in the final year of high school is mostly ineffective since it does not eliminate existing barriers to entry. However, earlier interventions of the benefits and costs of education to students and their parents may affect household behavior so that by the time children apply to college, both academic and financial barriers are less binding.

Table 1.1: Descriptive Statistics: Private, Public, and BHELPS schools

	Bogotá				BHELPS	
	Private schools		Public schools			
	Mean	(SD)	Mean	(SD)	Mean	(SD)
Panel A: Students						
Males	0.492	(0.500)	0.458	(0.498)	0.481	(0.500)
Age	17.648	(0.907)	17.641	(0.873)	17.655	(0.950)
Parent completed secondary	0.288	(0.453)	0.395	(0.489)	0.401	(0.490)
Parent completed higher education	0.580	(0.494)	0.156	(0.363)	0.183	(0.387)
Family income (<1 MW)	0.028	(0.165)	0.144	(0.351)	0.172	(0.377)
Family income (1-2 MWs)	0.246	(0.431)	0.559	(0.497)	0.467	(0.499)
Family income (>2 MWs)	0.726	(0.446)	0.297	(0.457)	0.361	(0.480)
Born in Bogotá					0.847	(0.360)
Internet at home					0.691	(0.462)
Victim of violence					0.035	(0.183)
Student works					0.170	(0.375)
Has repeated at least one grade					0.251	(0.434)
Risk averse					0.851	(0.357)
Perceived high academic ranking					0.410	(0.492)
Perceived High self-efficacy					0.352	(0.478)
Perceived high likelihood of enrollment					0.843	(0.364)
Panel B: Schools						
Number of students (2010-2012)	111.15	(168.48)	99.66	(48.08)	93.65	(40.66)
SABER 11 score (2010-2012)	0.874	(0.809)	0.117	(0.254)	0.139	(0.248)
Morning shift	0.191	(0.393)	0.547	(0.498)	0.633	(0.482)
Afternoon shift	0.019	(0.137)	0.390	(0.488)	0.348	(0.476)
Single shift	0.790	(0.407)	0.063	(0.243)	0.019	(0.138)
School has computer lab					0.964	(0.187)
Total number of students	37,068		37,787		6,636	
Total number of schools	790		570		116	

Source: Authors' calculations from ICFES and BHELPS survey.
Notes: Statistics for Bogotá are based on ICFES, which includes the universe of schools offering an academic track. Using date of birth, we compute each student's age on December 31, 2013. The number of students is the average number of individuals who sat for the SABER 11 exam in each year from 2010-2012. SABER 11 scores are standardized with respect to each year's national average. The difference between private and public schools is statistically significant at the 1% level for all attributes except age and number of students.

Table 1.2: Descriptive Statistics for the 2013 Cohort Test Scores and Enrollment Choices

	Bogotá				BHELPS	
	Private schools		Public schools			
	Mean	(SD)	Mean	(SD)	Mean	(SD)
Panel A: Exit Exam						
Overall Score	0.864	(1.192)	0.138	(0.841)	0.127	(0.823)
Math	0.708	(1.231)	0.046	(0.884)	0.024	(0.868)
Language	0.702	(1.060)	0.156	(0.870)	0.171	(0.864)
Panel B: College Enrollment						
Enrolled	0.571	(0.495)	0.426	(0.495)	0.438	(0.496)
Public College	0.147	(0.354)	0.278	(0.448)	0.287	(0.452)
Private College	0.424	(0.494)	0.148	(0.355)	0.151	(0.358)
Top-10 College	0.160	(0.366)	0.011	(0.106)	0.011	(0.102)
Academic degree (4-year)	0.370	(0.483)	0.098	(0.298)	0.092	(0.290)
Vocational degree (2-year)	0.201	(0.400)	0.328	(0.469)	0.346	(0.476)
STEM degree	0.211	(0.408)	0.054	(0.227)	0.049	(0.215)

Source: Authors' calculations from ICFES, SNIES, and BHELPS survey.
Notes: Statistics for Bogotá are based on 2013 ICFES and 2014 SNIES data, which includes the universe of schools offering an academic track. SABER 11 scores are standardized with respect to the 2013 national average. The difference between private and public schools is statistically significant at the 1% level for all attributes.

Table 1.3: Balance in Baseline Student and School Characteristics by Treatment

	Control		Treatment		Difference
	Mean	(SD)	Mean	(SD)	P-value
Panel A: Students					
Males	0.485	(0.500)	0.477	(0.500)	0.647
Age	17.641	(0.938)	17.668	(0.962)	0.469
Born in Bogotá	0.851	(0.357)	0.843	(0.364)	0.539
Parent completed secondary	0.405	(0.491)	0.396	(0.489)	0.567
Parent completed higher education	0.194	(0.396)	0.172	(0.377)	0.263
Family income (<1 MW)	0.167	(0.373)	0.176	(0.381)	0.585
Family income (1-2 MWs)	0.467	(0.499)	0.468	(0.499)	0.976
Family income (>2 MWs)	0.366	(0.482)	0.357	(0.479)	0.720
Internet at home	0.711	(0.453)	0.672	(0.470)	0.090
Victim of violence	0.034	(0.181)	0.035	(0.184)	0.816
Student works	0.163	(0.370)	0.176	(0.381)	0.329
Has repeated at least one grade	0.247	(0.431)	0.255	(0.436)	0.648
Risk averse	0.856	(0.351)	0.845	(0.362)	0.400
Perceived high academic ranking	0.425	(0.494)	0.395	(0.489)	0.111
Perceived high self-efficacy	0.349	(0.477)	0.355	(0.479)	0.714
Perceived high likelihood of enrollment	0.841	(0.365)	0.844	(0.363)	0.862
Panel B: Schools					
Number of students (2010-2012)	95.007	(48.106)	92.349	(31.826)	0.740
SABER 11 score (2010-2012)	0.160	(0.215)	0.118	(0.275)	0.379
Morning shift	0.641	(0.480)	0.625	(0.484)	0.867
Afternoon shift	0.337	(0.473)	0.359	(0.480)	0.807
Single shift	0.023	(0.149)	0.016	(0.125)	0.808
School has computer lab	0.970	(0.172)	0.958	(0.201)	0.741
Total number of students	3,259		3,377		
Total number of schools	59		57		

Source: Authors' calculations from ICFES and baseline BHELPS survey.
Notes: Using date of birth, we compute each student's age on December 31, 2013. The number of students is the average number of individuals who sat for the SABER 11 exam in each year from 2010-2012. SABER 11 scores are standardized with respect to each year's national average. The last column presents the p-value of the difference in the attribute between treatment and control groups calculated by regression with clustered standard errors at the school-level. A joint-test for balance rejects that individual and school-level characteristics explain the likelihood of attending a treatment school, with a p-value of 0.680.

Table 1.4: Balance in Baseline Student Knowledge and Beliefs by Treatment

	Control		Treatment		Difference
	Mean	(SD)	Mean	(SD)	P-value
Knows ICETEX	0.700	(0.458)	0.688	(0.463)	0.612
Knows FESBO	0.180	(0.384)	0.168	(0.374)	0.295
Premium Error: Vocational	0.696	(1.539)	0.615	(1.475)	0.105
Premium Error: Academic	1.184	(1.259)	1.100	(1.234)	0.091

Source: Authors' calculations from baseline BHELPS survey.
Notes: The last column presents the p-value of the difference in the attribute between treatment and control groups calculated by regression with clustered standard errors at the school-level.

Table 1.5: Baseline Characteristics by Direction of Belief Error

	Premium Error: Vocational			Premium Error: Academic		
	Under	Over	Difference P-value	Under	Over	Difference P-value
Panel A: Students						
Treatment group	0.529	0.505	0.332	0.542	0.504	0.132
Males	0.471	0.484	0.590	0.475	0.484	0.691
Age	17.765	17.624	0.001	17.800	17.621	0.000
Born in Bogotá	0.829	0.852	0.130	0.838	0.850	0.446
Parent completed secondary	0.389	0.402	0.485	0.390	0.402	0.539
Parent completed higher education	0.183	0.186	0.811	0.173	0.188	0.313
Family income (<1 MW)	0.229	0.160	0.000	0.211	0.164	0.006
Family income (1-2 MWs)	0.458	0.467	0.601	0.498	0.463	0.103
Family income (>2 MWs)	0.314	0.373	0.002	0.291	0.373	0.000
Internet at home	0.665	0.699	0.076	0.670	0.697	0.143
Victim of violence	0.040	0.034	0.445	0.033	0.034	0.837
Student works	0.194	0.168	0.092	0.186	0.169	0.298
Has repeated at least one grade	0.280	0.243	0.065	0.288	0.244	0.036
Risk averse	0.826	0.858	0.043	0.821	0.858	0.024
Perceived high academic ranking	0.357	0.421	0.003	0.321	0.422	0.000
Perceived high self-efficacy	0.370	0.347	0.242	0.337	0.350	0.533
Perceived high likelihood of enrollment	0.788	0.852	0.000	0.770	0.854	0.000
Panel B: Schools						
Number of students (2010-2012)	91.602	94.118	0.123	91.216	94.145	0.101
SABER 11 score (2010-2012)	0.133	0.143	0.408	0.119	0.145	0.038
Morning shift	0.620	0.637	0.417	0.634	0.635	0.971
Afternoon shift	0.359	0.343	0.441	0.355	0.344	0.615
Single shift	0.021	0.020	0.723	0.010	0.021	0.000
School has computer lab	0.971	0.964	0.260	0.961	0.965	0.374

Source: Authors' calculations from ICFES and baseline BHELPS survey.
Notes: The difference column presents the p-value of the difference in the attribute between students who over and under estimate earning premiums and are calculated by regression with clustered standard errors at the school-level.

Table 1.6: Treatment Effects on Knowledge and Beliefs

	Knows ICETEX (1)	Knows FESBO (2)	Premium Error: Vocational (3)	Premium Error: Academic (4)
Panel A: After, All students in follow-up				
Treat	0.039***	-0.001	0.02	-0.023
	(0.014)	(0.011)	(0.044)	(0.040)
Observations	6,003	5,799	5,920	5,913
Panel B: After, Matched with baseline				
Treat	0.048***	-0.001	0.015	-0.017
	(0.015)	(0.012)	(0.043)	(0.040)
Observations	5,427	5,242	5,361	5,355
Panel C: ANOVA				
Treat	0.047***	0.001	0.012	-0.005
	(0.013)	(0.011)	(0.044)	(0.039)
Observations	5,347	5,096	5,053	5,046
Panel D: Difference-in-differences				
Treat × Post	0.046**	0.007	0.077	0.043
	(0.018)	(0.014)	(0.062)	(0.054)
Post				
	0.125***	0.025**	-0.097**	-0.110***
	(0.011)	(0.010)	(0.049)	(0.042)
Observations	10,861	10,591	10,538	10,532
Mean(y) at baseline	0.694	0.174	0.655	1.141

Source: Authors' calculations from BHELPS survey.
Notes: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each column and panel correspond to a separate OLS regression. Panels A and B control for student and household attributes (male, age, age squared, family income, and parental education) and school characteristics (average SABER 11 score in previous years, has computer lab, shift indicators, and school size). Panel C presents coefficients for difference-in-difference regression that control for individual fixed-effects. Standard errors are clustered at school-level.

Table 1.7: Treatment Effects on Beliefs by Direction of Belief Error

	Premium error: Vocational		Premium error: Academic	
	Under	Over	Under	Over
	(1)	(2)	(3)	(4)
Panel A: After, Matched with baseline				
Treat	0.097	-0.009	0.008	-0.023
	(0.122)	(0.045)	(0.116)	(0.043)
Observations	601	4,452	515	4,531
Panel B: Difference-in-differences				
Treat × Post	0.107	0.052	-0.056	0.047
	(0.189)	(0.058)	(0.131)	(0.050)
Post	1.732***	-0.334***	1.498***	-0.289***
	(0.145)	(0.044)	(0.100)	(0.040)
Observations	1,236	8,993	1,060	9,162
Mean(y) at baseline	-1.604	0.974	-0.837	1.382

Source: Authors' calculations from BHELPS survey.
Notes: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each column and panel correspond to a separate OLS regression. Panels A controls for student and household attributes (male, age, age squared, family income, and parental education) and school characteristics (average SABER 11 score in previous years, has computer lab, shift indicators, and school size). Panel B presents coefficients for difference-in-difference regressions that control for individual fixed-effects. Standard errors are clustered at school-level.

Table 1.8: Treatment Effects on Test Scores

	Overall	Math	Language
	(1)	(2)	(3)
Panel A: All matched administrative			
Treat	-0.015	0.026	-0.013
	(0.029)	(0.034)	(0.028)
Observations	6,896	6,896	6,896
Panel B: Matched with baseline			
Treat	-0.012	0.032	-0.010
	(0.030)	(0.034)	(0.029)
Observations	6,309	6,309	6,309
Panel C: Matched with baseline and follow-up			
Treat	0.001	0.045	-0.003
	(0.033)	(0.038)	(0.032)
Observations	5,414	5,414	5,414

Source: Authors' calculations from ICFES and BHELPS survey.
Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each column and panel corresponds to a separate OLS regression that controls for student and household attributes (male, age, age squared, family income, and parental education) and school characteristics (average SABER 11 score in previous years, has computer lab, shift indicators, and school size). Standard errors are clustered at school-level.

Table 1.9: Treatment Effects on Enrollment Choices

	Enrolled College (1)	Private College (2)	Top-10 College (3)	Academic Degree (4)	STEM Degree (5)
Panel A: All matched administrative					
Treat	0.013	0.011	0.004	0.008	0.006
	(0.019)	(0.010)	(0.002)	(0.008)	(0.006)
Observations	6,868	6,868	6,868	6,868	6,868
Panel B: Matched with baseline					
Treat	0.012	0.015	0.005*	0.009	0.008
	(0.019)	(0.010)	(0.002)	(0.009)	(0.006)
Observations	6,289	6,289	6,289	6,289	6,289
Panel C: Matched with baseline and follow-up					
Treat	0.005	0.012	0.006**	0.011	0.009
	(0.020)	(0.011)	(0.003)	(0.009)	(0.006)
Observations	5,401	5,401	5,401	5,401	5,401
Mean(y) control group	0.444	0.153	0.011	0.095	0.05

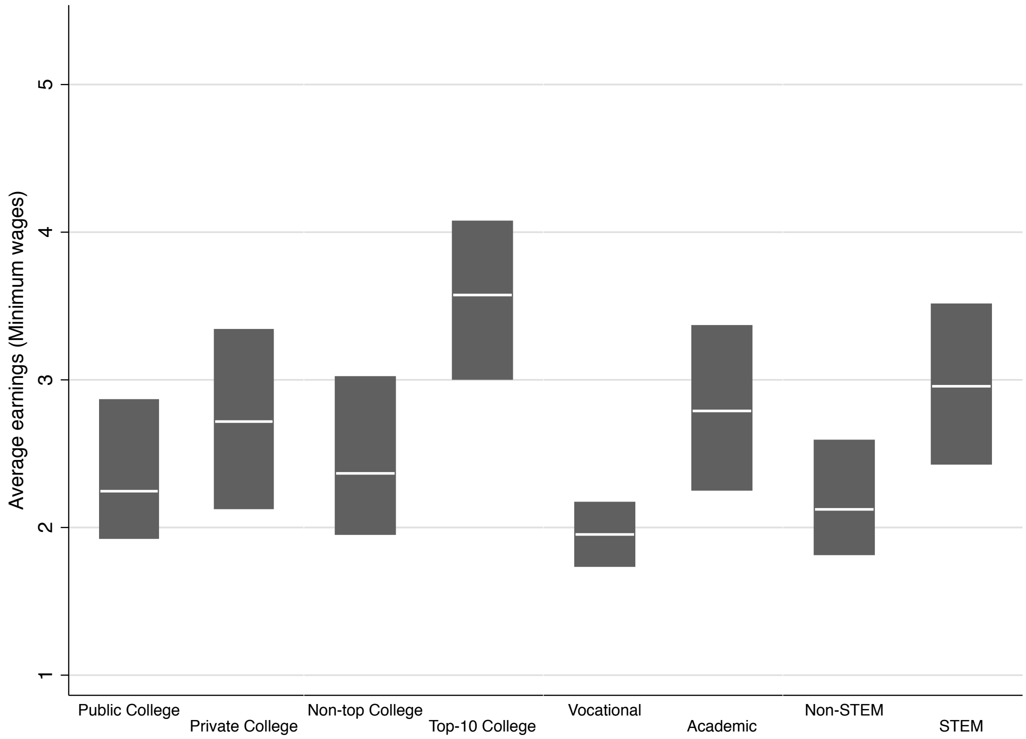
Source: Authors' calculations from ICFES, SNIES, and BHELPS survey.
Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each column and panel corresponds to a separate OLS regression that controls for student and household attributes (male, age, age squared, family income, and parental education) and school characteristics (average SABER 11 score in previous years, has computer lab, shift indicators, and school size). Standard errors are clustered at school-level.

Table 1.10: Treatment Effects by Family Income and Gender (baseline matched to administrative data)

	Knows ICETEX (1)	Overall score (2)	Math (3)	Language (4)	College Enrollment (5)	Private College (6)	Top-10 College (7)	Academic Degree (8)	STEM Degree (9)
Panel A: Treatment effects by family income									
Low income	0.061***	-0.045	0.008	-0.058	0.002	0.021**	0.001	0.007	0.008
	(0.018)	(0.036)	(0.036)	(0.035)	(0.021)	(0.009)	(0.002)	(0.008)	(0.006)
Middle income	0.028*	0.043	0.076	0.066	0.032	0.008	0.012**	0.014	0.009
	(0.015)	(0.044)	(0.046)	(0.042)	(0.025)	(0.019)	(0.005)	(0.017)	(0.013)
P-value (Low=Middle)	0.080	0.090	0.144	0.013	0.243	0.522	0.050	0.717	0.906
Observations	5427	6,309	6,309	6,309	6,289	6,289	6,289	6,289	6,289
Panel B: Treatment effects by Gender									
Female	0.029	-0.049	0.008	-0.054	-0.008	0.005	0.003	0.004	0.003
	(0.018)	(0.038)	(0.040)	(0.040)	(0.024)	(0.014)	(0.003)	(0.011)	(0.007)
Male	0.069***	0.026	0.057	0.035	0.033	0.024*	0.007*	0.014	0.014
	(0.019)	(0.039)	(0.042)	(0.035)	(0.022)	(0.013)	(0.004)	(0.012)	(0.010)
P-value (Female=Male)	0.069	0.109	0.289	0.063	0.122	0.265	0.421	0.491	0.336
Observations	5,427	6,309	6,309	6,309	6,289	6,289	6,289	6,289	6,289

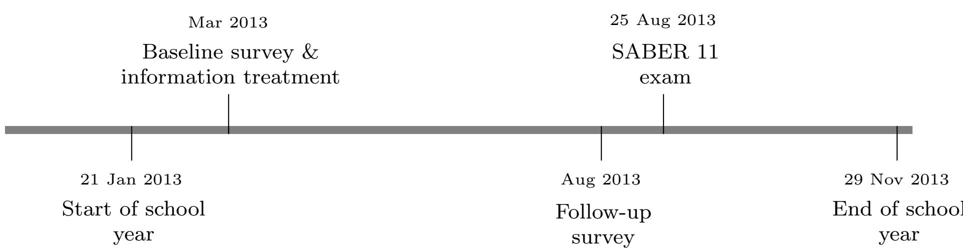
Source: Authors' calculations from ICFES, SNIES, and BHELPS survey.
Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each column and panel corresponds to a separate OLS regression that controls for student and household attributes (male, age, age squared, family income, and parental education) and school characteristics (average SABER 11 score in previous years, has computer lab, shift indicators, and school size). Standard errors are clustered at school-level.

Figure 1.1: Average Earnings of Recent Graduates



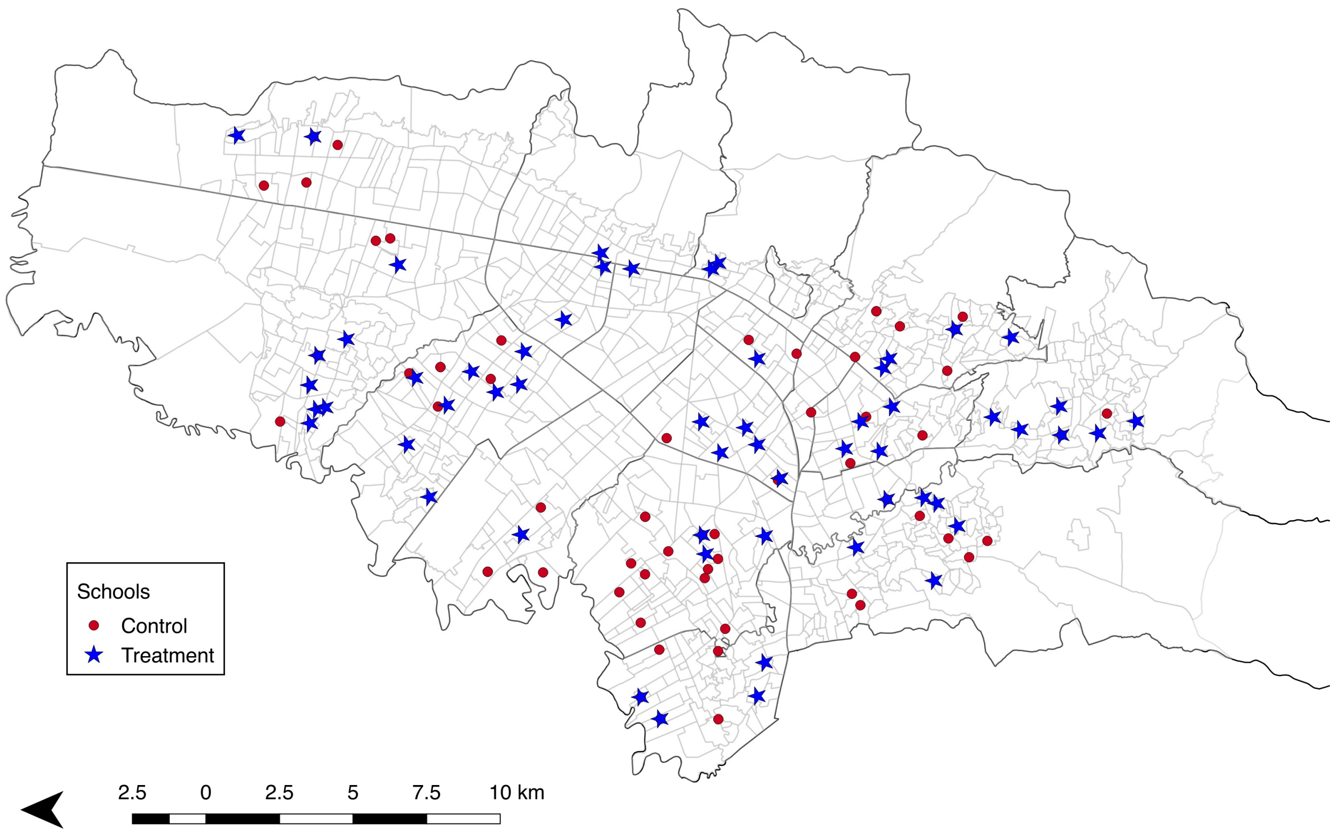
Source: Authors' elaboration from Labor Observatory data.
Notes: The figure shows the distribution of initial earnings for different categories of college and degree. Monthly earnings are expressed in minimum wages, and correspond to the average pay of recent graduates by college, level, and field as defined in Section 1.2. The grey box represents the 25th and 75th percentiles, the white line denotes the median.

Figure 1.2: Timing of Intervention and Data Recollection



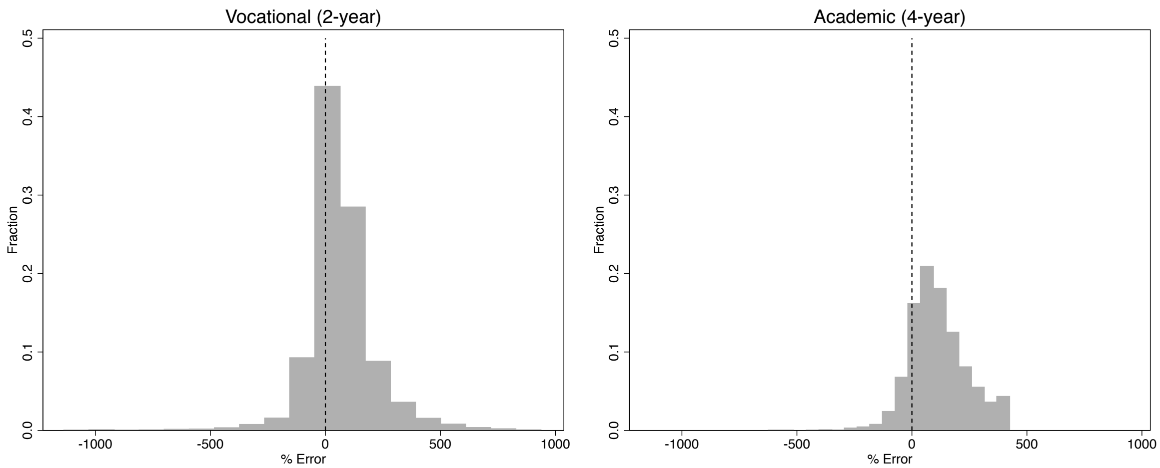
Source: Authors' elaboration.

Figure 1.3: Geographic distribution of treatment and control schools



Source: Authors' elaboration from Secretary of Education's School Census and BHELPS.

Figure 1.4: Distribution of Earning Premium Beliefs at Baseline



Source: Authors' elaboration from BHELPS baseline sample.
Notes: We calculate the error percentage as the difference between perceived and actual premiums divided by the actual premium. Let π^j denote the wage premium, with $j = \{\text{actual,perceived}\}$. Errors are calculated as $(\pi^{perceived} - \pi^{actual})/\pi^{actual}$.

Chapter 2

Do High School Peers Influence Post-Secondary Decisions? An Endogenous Network Approach

2.1 Introduction

While most of the peer effects literature in education has focused on academic performance and juvenile behavior, little is known about social influence on post-secondary decisions. This question is relevant for at least two reasons. First, conditional on student characteristics, college and major choices are key determinants of earnings (e.g. Dale and Krueger, 2002, 2011, Hoekstra, 2009, Hastings et al., 2013, Reyes et al., 2016). Second, post-secondary choices are determined by a number of factors beyond academic performance. For instance, Hoxby and Turner (2013) show that low-income high-achieving students apply to less selective colleges than their high-earning counterparts. Papay et al. (2015) show that performance labels influence college choice even though they provide no additional information. Zafar (2013) and Wiswall and Zafar (2015) find that heterogeneous preferences and tastes are the main determinants of major choice in college. Although it seems reasonable to believe that peers play a key role in post-secondary decisions, there is not enough evidence to support or reject this claim.

One of the main difficulties in addressing this question lies in the presence of three confounding sources of social influence: endogenous, exogenous and correlated effects. *Endogenous* effects correspond to the influence of peers' behavior on the individual behavior. They are the most relevant ones to researchers and policy makers because they reflect social spillovers, which have serious implications for school integration policies, and impact evaluation. *Exogenous* effects are the influence of the peers' characteristics on the individual behavior, and *Correlated* effects happen when individuals in the same group behave similarly because they share group characteristics. Most of the literature on peer effects and post-secondary choices is based on group interaction assumptions and is unable to separate endogenous from exogenous effects (Sacerdote, 2001, Fletcher, 2015, Luppino and Sander, 2015). A second body of literature exploits non-overlapping groups (De Giorgi et al., 2010), or social networks (Mora and Oreopoulos, 2011, Burgess et al., 2011) to identify endogenous peer effects. In those papers, the key assumption is that networks are formed exogenously, which is less realistic when it comes to social relationships.

This paper measures the influence of high school peers on post-secondary decisions. Endogenous peer effects are identified in a network framework. The data on social relationships and post-secondary aspirations was collected in 2013, with a survey conducted in 116 high schools from Bogotá, Colombia. The sample includes over 6,000 senior-year high-school students that are about to take the national exit exam. A year after graduation, the survey was matched to official administrative records to follow-up on exit exam scores and enrollment choices. In the benchmark social influence model, individuals choose based on their peers' choices and characteristics, and class fixed effects account for observed and unobserved class correlated factors. The exogenous network assumption is then relaxed using a selection-correction model.

The main results indicate small significant endogenous peer effects on some of the aspirations and exit exam outcomes, but not on the enrollment choices. Restricting the network to study mates or reciprocal nominations yields similar results. This evidence suggests that peers' influence on aspirations and academic performance fails to translate into actual enrollment choices. This can be explained by other factors that determine college enrollment, such as financial constraints, and by the fact that the peer effects on aspirations and test scores are relatively small. I also find that models that omit class correlated effects and social selection find positive and significant endogenous effects for all outcomes. This confirms the importance of controlling for these sources of bias, otherwise there is a big chance of overestimating the role of peers.

The paper contributes to the literature in a at least three ways. First, it estimates endogenous peer effects on post-secondary decisions while controlling for unobserved correlated effects and social selection. This is an improvement with respect to papers that are based on group interactions, which are unable to identify endogenous effects. Furthermore, results confirm that assuming exogenous relationships, as in Mora and Oreopoulos (2011) or Burgess et al. (2011), could lead to biased estimates. Second, this paper studies the effects of high school peers on post-secondary intensive and extensive margin decisions. Fletcher (2015) focus on college enrollment, but does not assess the social effects on the intensive margin. Sacerdote (2001) and De Giorgi et al. (2010) successfully exploit random assignment rules to deal with selection problems, however they are restricted to college peers and their effects on major choices. Moreover, having data on aspirations and actual enrollment allows studying the role of school peers at two different stages of the decision process. As will be seen, peers have more influence on aspirations than on actual enrollment choices. Third, it considers different types of social relationships, finding that results are robust to how network are defined.

The remainder of this chapter is organized as follows. The next section provides a brief summary of the literature, focusing on the identification of endogenous peer effects, and the existing evidence of peer effects on post-secondary decisions. Sections 3 and 4 describe the data and the empirical strategy. Section 5 presents the main results, and the last section concludes.

2.2 Previous literature

2.2.1 Identification of Endogenous Peer Effects

The major empirical challenge in the peer effect literature is to identify among three sources of social influence: *endogenous, exogenous* and *correlated effects*. Endogenous effects correspond to the influence of peers' behavior on the individual behavior. Exogenous effects capture the influence of the peers' characteristics on the individual behavior. Social effects are usually defined as the sum of these two effects. Correlated effects, on the other hand, happen when individuals in the same group behave similarly because they share group characteristics. There are at least two identification problems to consider. On the one hand, group selection and unobserved correlated effects (e.g. teachers quality) are potential sources of bias. On the other hand, identifying endogenous from exogenous effect in the linear-in-mean model requires prior information on the composition of the reference group. Manski (1993b) refers to this as the *reflection problem*.

The endogenous peer effects are the most interesting ones because they reflect social spillovers. Identifying them is highly relevant for researchers and policy makers for at least two reasons. First, in the presence of spillovers, school and class composition have direct effects on education outcomes. This has been a central argument to promote more integrated and socially mixed learning environments, and has motivated a whole literature that studies the effects of integration policies (e.g. Hoxby, 2000, Hanushek et al., 2002, Angrist and Lang, 2004). Second, social spillovers may partially account for the benefits of a program. Moreover, in scenarios where treatment and control groups are in contact, spillovers may seriously bias the estimated impacts (e.g. Kremer et al., 2009, Bobonis and Finan, 2009, Angelucci and De Giorgi, 2009).

Most of the literature on peer effects assumes group interactions, i.e. all members of a dorm, classroom or school are considered peers. The most relevant papers in this category address the selection problem and estimate social effects. Some of them exploit a random assignment rule (e.g. Sacerdote, 2001, Zimmerman, 2003, Duflo et al., 2011, Imberman et al., 2012, Deming et al., 2014, Griffith and Rask, 2014). Other papers use the idiosyncratic variations in composition within schools (e.g. Hoxby, 2000, Hanushek et al., 2003, Ding and Lehrer, 2007). The main limitation of both of these approaches is that they are unable to separate endogenous from exogenous effects.[1] Besides, it is possible that unobserved correlated factors bias the estimated social effects, especially when peer groups are defined at the class or school level. The only paper based on group interactions that identify endogenous effects is Lee (2007), with a methodology that takes advantage of the variations in group size.

When more detailed information on the social relationships is available, it is possible to relax the groups interaction assumption. In this case, individuals are no longer influenced by all the students in the group but only by those in each person's reference group. This can be done when students are affiliated to multiple overlapping groups (e.g. Laschever, 2009, De Giorgi et al., 2010) or when there is detailed information on social networks (e.g. Bramoullé et al., 2009, Calvó-Armengol et al., 2009, Lin, 2010, Fryer and Torelli, 2010). The main advantage of this approach is that it exploits the non-transitivity to identify endogenous from exogenous peer effects. Moreover, it is possible to capture unobserved correlated effects and rule out selection bias using group fixed-effects.

The main limitation of this approach is that it assumes that networks are formed exogenously. This assumption is unrealistic when it comes to social relationships. In fact,

[1] Sacerdote (2001) proposes a structural model to achieve this, but concludes that identification of endogenous effects is only possible under very restrictive assumptions.

it is violated if links are formed based on factors that are also related to post-secondary decisions. This may happen when students tend to associate with similar others (homophily), and common characteristics also determine post-secondary choices. It might also result from network dependency. For instance, if popularity is correlated with school performance or post-secondary decisions (e.g. Calvó-Armengol et al., 2009, Conti et al., 2013). Two recent papers have addressed this problem with selection-correction models in the spirit of Heckman (1979). Goldsmith-Pinkham and Imbens (2013) propose a network formation model where individuals are more likely to be linked when they have similar observed characteristics and belong to the same unobserved type. Hsieh and Lee (2014) follow a similar strategy, but generalize the selection step in at least two ways. First, their model considers any type of social relationships, including directed and undirected links. Second, it allows for n continuous unobserved characteristics that determine link formation.

2.2.2 Social Influence on Post-Secondary Decisions

Among the papers based on group interaction that studies the social influence on post-secondary choices, only three address in some way the selection bias problem. The best case for exogenous selection is Sacerdote (2001), who exploits the random assignment of dorms at Dartmouth College. Results show that roommates do influence academic performance and affiliation to social groups, but have no significant effect on major choice. Fletcher (2015) attenuates the bias by controlling for school fixed effects and instrumenting classmates behavior with their parents' expectations. The authors find that school classmates affect the probability of enrolling in college. Luppino and Sander (2015) also reduce the bias by controlling for campus and application-admissions patterns fixed effects. The main results indicate that attending a college campus with stronger peers in science negatively affects the probability of completing a Science, Technology, Engineering, and Mathematics (STEM) degree. A key limitation of these papers is that they are unable to separate endogenous from exogenous effects.

Three papers study social effects on post-secondary decisions following a network approach. De Giorgi et al. (2010) take advantage of randomly assigned overlapping classes at Bocconi University to map an exogenous network of peers. The authors find that first-year classmates do influence major choices (between economics and finance). Mora and Oreopoulos (2011) use data from Spain on self-reported friendship nominations to estimate the influence of high school peers on the intentions to drop out. Results indicate that friends have no significant effect. Similarly, Burgess et al. (2011) use high school friends networks from Bristol but focus on the exogenous effects. The main findings suggest that peers parents' characteristics have positive effects on post-secondary aspirations. The empirical strategy of these three papers relies on the exogeneity of the network. While this is a natural condition in De Giorgi et al. (2010), given the random assignment, Mora and Oreopoulos (2011) and Burgess et al. (2011) assume that friendship relationships are formed exogenously. As discussed in Section 2.2.1, this may be an unrealistic assumption.

This paper estimates the effect of high school peers on post-secondary decisions, relaxing the exogenous network assumption. It does so by using the Hsieh and Lee (2014) selection-correction model. The next section presents the survey conducted and the matched administrative records. The model and the identifying assumptions are described in section 3.3.3.

2.3 Data description

In order to study the influence of peers on post-secondary aspirations and enrollment choices, this paper conducted a survey in 116 public high schools from Bogotá, Colombia in August 2013, one week before taking the standardized national exit exam.[2] I focus on public schools because they have significantly lower enrollment rates than private schools. At most two senior-year classes were randomly selected in each school, for a total of 203 classes and over 6,000 students. Individuals were asked to list the closest peers, and state their post-secondary aspirations. I match the individual students to official records from the Ministry of Education to follow-up on exit exam scores and college enrollment in 2014. Both the survey questionnaires and the administrative records provide information on demographics and socioeconomic status.

Panel A of Table 2.1 presents the descriptive statistics of the students' characteristics. The students' average age is 17.64 and males are slightly underrepresented. Most of the students come from a low socioeconomic background. In fact, only 55% have a parent who completed secondary, 16.4% of which have a post-secondary degree. Moreover, less than 32% of the families have income higher than 2 minimum wages, and can be considered

[2] This project was reviewed and approved in advance by the Institutional Review Board for the protection of human subjects of the University of Illinois at Urbana-Champaign (IRB #13570).

non-poor.[3] Consistently, 18% of the students report to be working before graduating from high school.

This paper estimates the influence of peers on a broad set of post-secondary outcomes. To measure aspirations, students are first asked whether they want to pursue higher education; 98.9% of them do. They are then asked to choose the college and degree they are most interested in. As can be seen in Panel B of Table 2.1, 23.7% and 45.1% of the students aspire to private colleges and top-10 colleges respectively.[4] There are two degree levels, vocational (2 years) and academic (4 years). Over 85% of the students in the sample aim for the an academic degree, and 40.8% are interested in STEM degrees.[5] Notice that in Colombia students apply to specific degrees with relatively few chances to change afterward. Using administrative records on entry wages, an expected monetary value is assigned to each college-degree choice.[6] Students aspire to earn on average 2.56 minimum wages. This measure captures better the heterogeneity of the higher education system. In fact, there is a large variation in expected wages by college and degree and most of the

[3] Wages are expressed in minimum monthly wages, which is a commonly used measure in Colombia. The 2013 monthly minimum wage was 535,600 Colombian Pesos, equivalent to 288 US Dollars. The official poverty line for a family of four in Bogotá is 1.7 minimum wages.

[4] According to the 2012 Higher education exit exams (SABER PRO), the top-10 colleges in Colombia are (in order): *Universidad de los Andes*, *Universidad Nacional* (Bogotá), *Universidad del Rosario*, *Universidad Externado*, *Universidad Icesi* (Cali), *Universidad Eafit* (Medellín), *Universidad de la Sabana*, *Universidad Javeriana*, *Universidad Nacional* (Medellín), and *Universidad del Norte* (Barranquilla). *Universidad Nacional* (Bogotá and Medellín) are the only public Universities ranked top-10.

[5] I classify as STEM degrees all academic degrees from Agronomy, animal sciences, veterinary medicine, medicine, bacteriology, biology, physics, mathematics, chemistry, geology, business, accounting, economics, and all engineering fields.

[6] Entry wages are defined as the 2011 average monthly wages of 2008-2011 graduates by field of study, degree level and college. When there are no observations in a particular subcategory, average wages by field and degree level are used instead. These estimates are based on data from the Labor Observatory of the Ministry of Education, which previously linked higher-education graduates to social security records. The expected wage for students who are not interested in post-secondary education, or fail to report a field or a degree level, is set to one minimum wage.

current literature does not take this into consideration. For instance, in Colombia, the average entry wage premiums for top-10 colleges, academic degrees, and STEM degrees are 1.05, 1 and 0.85 minimum wages, respectively.

Exit exam scores and college enrollment (in 2014) are obtained by matching the survey to administrative records. [7] Descriptive statistics of enrollment choices and exit exam scores are presented in Panel C and D of Table 2.1. Students enrollment outcomes are far below their aspirations. Only 43.9% enroll, and 1.1% make it to a top-10 college. Moreover, only 9.2% and 4.9% of the students opt for academic (4 years) and STEM degrees. As a result of this, the expected wage of the enrollment choice falls to 1.47 minimum wages. One of the main reasons why only a fraction of the students achieves their post-secondary aspirations is that they face major financial constraints.[8] As for the exit exam, the students in the sample perform slightly better than the national average, particularly in language and social sciences.

In order to map social networks, students are asked to nominate the three classmates they spend most time with.[9] Each classroom is considered as a separate network. A

7 The Colombian Institute for the Promotion of Higher Education -ICFES- records provide scores for the high school exit exam, and allow following up the students in the National Information System for Higher Education -SNIES. The matching rates for ICFES and ICFES-SNIES are 98.6% and 98.4%, respectively, and there are no significant differences between matched and unmatched students. Exit exam scores are normalized with respect to the national population. The national exit exam (SABER 11) is a requirement for college application. The overall score is a weighted average of the following subjects: Mathematics (3), language (3), sciences (3), social sciences (2) and philosophy (1).

8 Vocational programs (2-year) are for the most part free, but academic careers (4-year) are not. In public universities tuition fees are proportional to the family income, however acceptance rates are very low. High-quality private universities are expensive, and there are very few scholarships for low-income students. Funding programs are available, however they require a co-debtor, a restriction that is binding to low-income students. Consistently, 64% of the students in the sample believe that financial constraints are the main barrier to higher education.

9 There were no specific instructions regarding the ordering, therefore there are no reasons to believe

network example is presented in Figure 2.1. Nodes represent students, and each directed edges is a nomination. The color and diameter of the nodes correspond to the students' aspired wages, and the number of received nominations (in-degree), respectively. As can be seen, most of the students nominate three peers and there are many reciprocal nominations. Moreover, in most cases social relationships are not transitive, i.e. students have peers that are not connected to each other. As will be seen in the next section, this is a key condition to identify endogenous peer effects. Furthermore, students with low aspirations nominate peers with high aspirations and vice versa. Identifying the extent to which these students are influencing each other is the main purpose of this paper.

Network statistics are presented in Panel E of Table 2.1. A total of 5,909 students nominate at least one nominated peer in the sample, in 203 classrooms .The average size of the classrooms is 29.11. The average student nominate 2.65 peers, for a total of 15,684 directed links. Students are also asked about the kind of activities they do with each one of their nominees. It seems reasonable to believe that study mates are more influential than regular peers when it comes to post-secondary decisions. To test this, the network is restricted to peers who study together. In total, 5,590 students nominate at least one study mate, with an average of 2.48 study mates. I also test whether endogenous effects are larger when the network is restricted to reciprocal nominations. In this case, there are 5,498 links, and 5,438 students, with an average of 2.02 reciprocal nominations. All estimations presented in section 3.4 are based on sub-samples that iteratively remove miss-

that the first nomination is in some way more important than the others. While students had the option to nominate less peers, most of them listed three.

ing observations (on both covariates and outcomes) and isolated nodes, i.e. students who have no nominations left in the sub-sample. Networks with less than 15 students are also dropped to guarantee sufficiently large networks, with enough within-group variation. The exact number of networks and students used in each set of regressions is presented in the respective table.

2.4 Empirical Strategy

When network data is available, social influence is usually modeled with spatial econometric methods. In this context, the weighting matrix W is defined by social interactions. In directed networks, W is asymmetric and the element (i, j) is 1 if student i nominates student j and 0 otherwise. When the network is undirected, W is symmetric. This section briefly describes the Spatial Autoregressive Model (SAR) and the conditions under which endogenous peer effects are identified. It then presents the Hsieh and Lee (2014) Selection-correction SAR model (SCSAR), an endogenous network approach that accounts for social selection bias.

2.4.1 Spatial Autoregressive Model (SAR)

In the benchmark Spatial Autoregressive Model model (SAR), presented in Equation 2.1, the outcome Y_g of individuals belonging to group g (of size n_g) depends on the average outcome of her peers ($W_g Y_g$), her characteristics (X_g) and the average characteristics of her peers (W_gX_g). It is also possible to control for group fixed effects (α_g).

$$Y_g = \lambda W_g Y_g + X_g\beta_1 + W_g X_g \beta_2 + l_g\alpha_g + \epsilon_g, \quad \epsilon_g \sim N(0, \sigma_\epsilon^2 I_{n_g}), \quad g = 1, ..., G \tag{2.1}$$

Bramoullé et al. (2009) proves that endogenous and exogenous effects, represented by λ and β_2, are identified if and only if I, W and W^2 are linearly independent.[10] This condition is satisfied when: i. There are endogenous and/or exogenous effect and they don't cancel out; ii. social relationships are not transitive, i.e. students nominate peers who are not necessarily connected to each other. Note that the non-transitivity condition is violated when individuals interact in groups.[11] Moreover, since none of the identification conditions depend on the symmetry of the social interaction matrix, these results are valid for both directed and undirected (reciprocal) networks.

The group fixed-effects (α_g) account for observed and unobserved factors that are common to students, such as teacher quality, school environment and infrastructure. Fixed effects also control for non-observed factors that determine school and class selection. In the presence of group fixed effects, endogenous effects are identified as long as I, W, W^2 and W^3 are linearly independent, a condition that is very unlikely to be violated in sufficiently large networks (Bramoullé et al., 2009).

[10] In this context, W^2 is a matrix that characterizes peers' peers: the off-diagonal element (i, j) is equivalent to 1 if student i peers' nominate student j, and 0 otherwise.

[11] A weaker version of this proposition states that, even when the network is transitive, effects are still identified as long as students don't interact in groups.

2.4.2 Selection-correction SAR model (SCSAR)

One of the main limitations of the SAR model is that it assumes that networks are formed exogenously. As discussed in section 2.2, this can be an unrealistic assumption when it comes to social relationships. Hsieh and Lee (2014) address this problem using a two-stage procedure that explicitly models the network formation process and corrects for potential selection bias.

In this framework, the probability that student i nominates student j (from the same group g) is determined by the distance between them in observed (C_g) and unobserved (Z_g) characteristics. More specifically, the authors model directed dyads (nominations) using the following logistic regressions:

$$P(w_{ij,g}|C_g, Z_g) = \Lambda(\delta_0 + \sum_{q=1}^{\bar{q}} \delta_{1q}|c_{iq,g} - c_{jq,g}| + \sum_{d=1}^{\bar{d}} \gamma_d|z_{id,g} - z_{jd,g}|), \quad i, j \in g, g = 1, ..., G \tag{2.2}$$

Where $|c_{iq,g} - c_{jq,g}|$ is the distance in the q^{th} observed characteristic in C_g, and $|c_{iq,g} - c_{jq,g}|$ is the distance in the d^{th} unobserved characteristic in Z_g. The unobserved characteristics are represented by individual coordinates in a $\bar{d}$-dimension latent space, which are estimated along with the other parameters. This model accounts for *homophily* based on both observed and unobserved characteristics. It also controls for *transitivity* -individual who have friends in common are more likely to be connected- which captures a significant part of the network dependency.

The second stage of the Selection-Correction SAR model (SCSAR) introduces a correction term that accounts for the conditional correlation between the unobservable factor that determine social selection (Z_g) and the SAR error term (ϵ_g) from equation 2.1. Assuming that these terms follow a joint normal distribution (with correlation matrix $\sigma_{\epsilon Z}$), the model can be written as follows:

$$Y_g = \lambda W_g Y_g + X_g \beta_1 + W_g X_g \beta_2 + l_g \alpha_g + Z_g \Sigma_Z^{-1} \sigma_{Z\epsilon} + u_g, \quad u_g \sim N(0, \sigma_u^2 I_{n_g}), g = 1, ..., G \quad (2.3)$$

Where the error term is such that $\sigma_u^2 = (\sigma_\epsilon^2 - \sigma_{\epsilon Z} \Sigma_Z^{-1} \sigma_{Z\epsilon})$. Note that when the all the correlations of $Z_{i,g}$ and $\epsilon_{i,g}$ are zero, i.e. social selection is independent of the second stage, the selection bias is irrelevant, and peer effects can be estimated using the benchmark SAR model. Since the differences in observed ($|c_{iq,g} - c_{jq,g}|$) and unobserved ($|c_{iq,g} - c_{jq,g}|$) characteristics are excluded from the second stage, the exclusion restriction is satisfied and the model is identified.

Both SAR and SCSAR models are estimated with MCMC Bayesian methods, following closely Hsieh and Lee (2014). In the case of the $SCSAR$ model, the parameters of interest are $(\lambda, \beta, \delta, \gamma, \sigma_\epsilon^2, \sigma_{\epsilon Z}, \{\alpha_g\})$ and the individual latent space coordinates $\{Z_g\}$ from Equations 2.2 and 2.3. For simplicity, the authors assume that the different dimensions of Z_g are linearly independent. They also normalize its correlation matrix ($\Sigma_Z = I_{\bar{d}}$), and force the correlations between ϵ and Z to be non-negative ($\sigma_{\epsilon Z} \geq 0$). In presence of multiple unobservables, the first dimensions are assumed to be more relevant to link formation

($|\delta_1| \geq ... \geq |\delta_{\bar{d}}|$). Based on Hsieh and Lee (2014), who find that two unobserved factors are enough to capture most of the selection-bias, I restrict the latent space to a maximum of two dimension. All the results presented in the next section are based on 20,000 iterations, with 5,000 burn-in steps.

2.5 Results

Table 2.2 presents the peer effects estimations on college enrollment when all nominated peers are considered. The table reports the posterior mean and the 95% confidence interval of each estimated coefficient. The first column corresponds to the SAR model without fixed-effects. The second column introduces class fixed-effects, which account for school/class selectivity, and observed and unobserved correlated factors. The third and fourth columns correspond to the SCSAR models with one and two unobserved characteristics, respectively. The vector X includes all the students characteristics described in section 2.3. The network formation model includes the difference between students in four observable characteristics: gender, age, parents' education and family income.[12]

The first thing to be noticed is that endogenous effects are smaller in models that account for class correlated effects and social selection. In fact, the SAR models without fixed effects (column 1) find significant endogenous effects of 5.1%, while the SAR model with fixed effects (column 2) and the SCSAR models (columns 3 and 4) estimates oscillate

[12] Students are considered the same age if their birthday is less than 6 months apart. They have the same parents' education or family income if they are classified in the same category (as defined in section 2.3).

between -0.6% and 0.4% and are statistically insignificant. It is worth noting that this is not because the estimates are less precise, the confidence intervals size is similar across models. As for the exogenous peer effects, results indicate that having peers who work reduces the probability of enrolling by 3.5%. The exogenous effect of age is borderline significant but economically small.

Student observed characteristics do affect enrollment and also determine the nominations. Students from more educated and wealthier parents, and who are not working, are more likely to enroll. As for link formation, results indicate that students are more likely to nominate same gender students, with similar parents' education but different income group. The differences in latent space coordinates are positive and significant in both models. So are the correlations between ϵ and the elements of Z_g. These results indicate that unobserved factors determine nominations, which further justifies using of a selection-correction model.

Table 2.3 reports the estimated endogenous peer effects on a broader set of post-secondary enrollment outcomes, including the students aspirations. These measures, described in section 2.3, allow studying social influence at two different stages of the decision process, and also capture the heterogeneity of the higher education system. The table also presents the estimated endogenous effects on exit exam scores. Although measuring peer effects on academic performance is not the main purpose of the paper, these results are interesting for two reasons. First, test scores are a key determinant of college enrollment, and could be a mechanism through which peers affects these choices. Second, most of the specialized literature has focused on academic performance, and results are therefore comparable.

There are two regularities that are worth mentioning. First, the SAR model without fixed effects always finds positive and significant peer effects, oscillating between 2% and 6.2% for post-secondary aspirations or choices, and up to 7.7% for test scores (column 1). Second, the estimated endogenous effects are much smaller when the models account for class fixed effects and social selection. The average estimated endogenous effect drop from 4.5% in the SAR, to 1.5% in the SAR(FE), and zero in the SCSAR(FE,$\bar{d} = 2$). Once again, this has nothing to do with the precision of the estimates; the confidence interval sizes oscillate around 2% for most of the outcomes, and are similar across models. These results confirm that unobserved correlated effects and social selection are seriously biasing the benchmark SAR estimates.

The most restrictive model (SCSAR(FE,$\bar{d} = 2$)) show that peers' influence is not homogeneous across outcomes. Endogenous effects are found to be positive and significant for only one of the aspirations outcomes; top-10 colleges with an estimated effect of 2.9%. However, there are no significant effects on actual enrollment choices. Peers also influence academic performance, with positive and significant effects on language (1.9%), sciences (1.7%), social sciences (1.8%) and overall score (2.9%). The estimated effects are similar in magnitude to those in Hsieh and Lee (2014), who find endogenous effects on GPA that oscillate between 1.9% and 2.9%, depending on the number of unobserved characteristics. Other papers that estimate peer effects on academic performance using selection-correction models, like Goldsmith-Pinkham and Imbens (2013) and Del Bello et al. (2015), find slightly larger effects. Note however that their models are not entirely comparable.[13]

Table 2.4 presents the estimated endogenous effects for networks that are restricted to study mates. Larger endogenous effects would indicate that study mates are more influential than regular peers when it comes to post-secondary choices. SAR and SAR(FE) models tend to find slightly larger effects for study mates, but the differences are relatively small. As for the most restrictive selection-correction models, some outcomes have positive and significant effects, such as aspire to enroll in a private and top-10 college, and overall score, sciences and social sciences, with estimated coefficients oscillating between 1.5% and 3.5%. Note that the effects are similar in magnitude to those in Table 2.3, which implies that in this context study mates are not more influential than close peers. This is not a surprising result given the study design. In fact, nominations are limited to the classrooms, and 95% of students report to study with their nominated peers.

Results are fairly similar when the network is restricted to reciprocal nominations. As can be seen in Table 2.5, the SAR models still find relatively large endogenous effects for all outcomes, while the selection-correction estimates are much smaller and for the most statistically insignificant. In this case, the SCSAR(FE,$\bar{d} = 2$) model finds that peers have a positive and significant effect on three aspirations outcomes (top-10 college, academic

[13] As seen before, Goldsmith-Pinkham and Imbens (2013) use a simpler selection step that considers only undirected links and a binary uni-dimensional unobserved characteristic. The estimated effects range between 9% and 15% depending on the models specification. Del Bello et al. (2015) use a selection correction model with one unobserved characteristic, and study the differential effect of friends who are also neighbors. The authors find that neighbors do not significantly influence each other, but non-neighbors do, with an estimated effect of 7.1%.

and STEM degree), two exit exam scores (overall and social sciences) and enrollment in an academic degree. Note that this is the only specification where endogenous peer effects on enrollment choices are statistically significant. Since in most cases the estimated effects are similar in magnitude to those found with all nominations, I conclude that the main results of the paper are robust to different specifications of the social networks.

2.6 Conclusions

This paper provides new evidence on the effect of high school peers on post-secondary decisions. It addresses this question using social network methods that exploit social networks to identify endogenous peer effects. A survey conducted on over 6,000 high school students from Colombia and matched to administrative records, provide detailed information on close peer nominations and post-secondary decisions and exit exam scores. The exogenous network assumption is relaxed using a selection-correction approach that explicitly model link formation.

The main results indicate that peers have a small influence on aspirations and academic performance, but not on enrollment choices. In fact, the selection-correction models find endogenous peer effects that are significant for the aspirations to enroll in a top-10 college, and two exit exam scores (social sciences and overall score), but not for the actual enrollment choices. Estimates based on study mates or reciprocal nominations yield similar results, confirming that these findings are robust to how networks are defined.

There are at least two reasons why peer effects on aspirations and test scores are not transmitted to post-secondary decisions. First, aspirations and academic performance are not the only determinants of college and major choices. This is particularly true in this context, where only a fraction of the students achieve their aspirations, and financial constraints are the main barrier to higher education. Second, the estimated endogenous effects on aspirations and exit exam scores are relatively small. The estimated coefficients of the effects that are statistically significant oscillate between 1.9% and 2.9%.

The results of this paper also confirm that models that omit class correlated effects and social selection are likely to overestimate the role of peers. In fact, the benchmark spatial autoregressive model finds large and significant endogenous effects for all outcomes, while the models that account for class fixed effects and social selection find much smaller effects. It is worth noting that this is not due to precision losses; the size of the confidence intervals is relatively similar across models.

Even in a context where financial barriers are less binding, it would probably take larger peer effects on aspirations and academic performance to find evidence of spillovers on post-secondary choices. Future research could address this questions by focusing on more affluent students, or countries with better access to higher education. The results might also change when a larger set of peers or role models is considered. While most of the specialized literature has focused on peers within school, it seems reasonable to believe that students are also influenced by non-school friends. It could also be the case that they are influenced by students they respect and admire, but with whom they have no social ties.

Table 2.1: Descriptive Statistics

	Obs.	Mean	Std. Dev.
A. Students Characteristics			
Male	6,128	0.481	0.500
Age	6,131	17.660	0.939
At least one parent completed secondary	5,975	0.392	0.488
At least one parent completed higher edu.	5,975	0.164	0.370
High family income	6,014	0.313	0.464
Student is working	6,076	0.180	0.385
B. Post-secondary aspirations			
Enrollment	6,131	0.989	0.105
Private College	6,131	0.237	0.425
Top-10 College	6,131	0.451	0.498
Academic (4 years) Degree	6,131	0.855	0.352
STEM Degree	6,131	0.408	0.491
Expected wage	6,131	2.565	1.100
C. Post-secondary enrollment			
Enrolls	5,993	0.439	0.496
Private College	5,993	0.151	0.358
Top-10 College	5,993	0.011	0.102
Academic (4 years) Degree	5,993	0.092	0.290
STEM Degree	5,993	0.049	0.215
Expected wage	5,993	1.476	0.728
D. Exit exam			
Average score	6,014	0.127	0.823
Mathematics	6,014	0.024	0.868
Language	6,014	0.171	0.864
Sciences	6,014	0.116	0.856
Social sciences	6,014	0.162	0.889
E. Social networks			
Classroom (network) size	203	29.108	5.731
Nominations (per student)	5,909	2.654	0.595
Studymates (per student)	5,590	2.482	0.706
Reciprocal nominations (per student)	5,438	2.022	0.778

Note: This Table presents the descriptive statistics of the full sample of students. Expected wages are expressed in minimum wages. Exit exam scores are normalized with mean zero and standard deviation equal to one with respect to the national population. Classroom (network) size and nominations statistics are based on students who list at least one friend in the sample. Study mates restrict the sample to students who list at least one study mate in the sample. Reciprocal nominations per student consider only students who have at least one reciprocal nominations in the sample.

Table 2.2: Peer Effects on Enrollment (All nominations)

	SAR (1)	SAR (FE) (2)	SCSAR (FE, $\bar{d}=1$) (3)	SCSAR (FE, $\bar{d}=2$) (4)
Endogenous Effects (λ)	0.051	0.004	0.002	-0.006
	[0.04, 0.06]	[-0.01, 0.02]	[-0.01, 0.02]	[-0.02, 0.01]
Peers Characteristics (Exogenous Effects β_2)				
Male	0.003	0.007	0.008	0.010
	[-0.01, 0.02]	[-0.01, 0.02]	[-0.01, 0.02]	[-0.01, 0.03]
Age	0.002	0.003	0.002	0.002
	[0.00, 0.00]	[0.00, 0.00]	[0.00, 0.00]	[0.00, 0.00]
At least one parent completed secondary	-0.002	-0.005	-0.007	-0.004
	[-0.02, 0.02]	[-0.02, 0.01]	[-0.03, 0.01]	[-0.02, 0.01]
At least one parent completed higher edu.	0.006	0.002	-0.003	0.001
	[-0.02, 0.03]	[-0.02, 0.03]	[-0.03, 0.02]	[-0.02, 0.03]
High family income	0.011	0.017	0.017	0.015
	[-0.01, 0.03]	[-0.00, 0.03]	[-0.00, 0.03]	[-0.00, 0.03]
Student is working	-0.019	-0.038	-0.037	-0.035
	[-0.04, -0.00]	[-0.06, -0.02]	[-0.06, -0.02]	[-0.06, -0.01]
Own Characteristics (β_1)				
Male	-0.008	-0.003	0.002	0.002
	[-0.04, 0.03]	[-0.04, 0.03]	[-0.03, 0.03]	[-0.03, 0.03]
Age	0.012	-0.012	-0.091	-0.090
	[0.01, 0.01]	[-0.02, -0.00]	[-0.11, -0.08]	[-0.10, -0.08]
At least one parent completed secondary	0.084	0.073	0.046	0.048
	[0.06, 0.11]	[0.04, 0.10]	[0.02, 0.07]	[0.02, 0.08]
At least one parent completed higher edu.	0.128	0.113	0.084	0.088
	[0.09, 0.17]	[0.07, 0.15]	[0.05, 0.12]	[0.05, 0.13]
High family income	0.071	0.074	0.065	0.063
	[0.04, 0.1]	[0.05, 0.10]	[0.04, 0.09]	[0.03, 0.09]
Student is working	-0.065	-0.075	-0.059	-0.058
	[-0.1, -0.03]	[-0.11, -0.04]	[-0.09, -0.03]	[-0.09, -0.02]
Social Selection (δ,γ)				
Intercept			-0.132	5.196
			[-0.36, -0.05]	[5.09, 5.61]
Same gender			1.732	0.487
			[1.59, 1.80]	[0.38, 0.53]
Same age			0.394	-0.049
			[-0.07, 1.22]	[-0.08, 0.01]
Same parents' education			-0.142	0.160
			[-0.99, 0.15]	[0.03, 0.21]
Same family income			0.219	-1.375
			[-0.04, 0.29]	[-1.44, -1.26]
$\lvert Z_{i1} - Z_{j1}\rvert$			7.064	4.744
			[4.60, 7.98]	[4.67, 5.17]
$\lvert Z_{i2} - Z_{j2}\rvert$				5.672
				[5.59, 5.91]
Error terms				
σ^2_ϵ	0.239	0.226	0.221	0.220
	[0.23, 0.25]	[0.22, 0.23]	[0.21, 0.23]	[0.21, 0.23]
$\sigma_{\epsilon Z_1}$			0.009	0.057
			[0.00, 0.02]	[0.04, 0.07]
$\sigma_{\epsilon Z_2}$				0.013
				[0.00, 0.03]

Note: Results are based on a sub-sample of 201 networks and 5,655 students that include all students with at least one nomination in the sample. Each column represents a separate Bayesian estimations based on 20,000 replications with 5,000 burn-in steps. Posterior means and 95% confidence intervals (in brackets) of the estimated parameters are reported.

Table 2.3: Endogenous Peer Effects on Post-secondary Decisions (All nominations)

	SAR (1)	SAR (FE) (2)	SCSAR (FE, $\bar{d}=1$) (3)	SCSAR (FE, $\bar{d}=2$) (4)
Post-secondary Aspirations				
Enroll	0.032	0.009	0.004	-0.026
	[0.01, 0.04]	[-0.00, 0.01]	[-0.01, 0.01]	[-0.07, -0.01]
Private college	0.050	0.020	0.019	0.011
	[0.04, 0.06]	[0.01, 0.03]	[0.01, 0.03]	[-0.00, 0.03]
Top-10 college	0.062	0.032	0.031	0.029
	[0.05, 0.07]	[0.02, 0.04]	[0.02, 0.04]	[0.02, 0.04]
Academic (4 year) degree	0.048	0.020	0.018	0.010
	[0.04, 0.06]	[0.01, 0.03]	[0.00, 0.03]	[-0.00, 0.02]
STEM degree	0.037	0.018	0.016	0.006
	[0.02, 0.05]	[0.00, 0.03]	[0.00, 0.03]	[-0.01, 0.02]
Expected wage	0.036	0.013	0.006	0.003
	[0.02, 0.05]	[-0.00, 0.03]	[-0.01, 0.02]	[-0.01, 0.02]
Enrollment Choices				
Enroll	0.051	0.004	0.002	-0.006
	[0.04, 0.06]	[-0.01, 0.02]	[-0.01, 0.02]	[-0.02, 0.01]
Private college	0.020	-0.007	-0.008	-0.011
	[0.01, 0.03]	[-0.02, 0.01]	[-0.02, 0.01]	[-0.02, 0.00]
Top-10 college	0.022	-0.001	-0.001	-0.004
	[0.01, 0.03]	[-0.01, 0.01]	[-0.01, 0.01]	[-0.02, 0.01]
Academic (4 year) degree	0.047	0.018	0.017	0.003
	[0.04, 0.06]	[0.01, 0.03]	[0.01, 0.03]	[-0.01, 0.02]
STEM degree	0.034	0.010	0.009	0.008
	[0.02, 0.05]	[-0.00, 0.02]	[-0.00, 0.02]	[-0.00, 0.02]
Expected wage	0.046	0.010	0.005	0.007
	[0.03, 0.06]	[-0.00, 0.02]	[-0.01, 0.02]	[-0.01, 0.02]
Exit Exam				
Overall score	0.077	0.040	0.034	0.029
	[0.07, 0.09]	[0.03, 0.05]	[0.02, 0.05]	[0.02, 0.04]
Mathematics	0.046	0.007	0.000	0.002
	[0.03, 0.06]	[-0.01, 0.02]	[-0.01, 0.01]	[-0.01, 0.02]
Language	0.052	0.025	0.020	0.019
	[0.04, 0.06]	[0.01, 0.04]	[0.01, 0.03]	[0.01, 0.03]
Sciences	0.069	0.034	0.027	0.017
	[0.06, 0.08]	[0.02, 0.05]	[0.02, 0.04]	[0.00, 0.03]
Social sciences	0.062	0.025	0.020	0.018
	[0.05, 0.07]	[0.01, 0.04]	[0.01, 0.03]	[0.01, 0.03]

Note: Results are based on a sub-sample of 201 networks and 5,655 students that include all students with at least one nomination in the sample. Each cell represents a separate Bayesian estimations based on 20,000 replications with 5,000 burn-in steps. Posterior means and 95% confidence intervals (in brackets) of the endogenous peer effects are reported. Expected wages are expressed in minimum wages and exit exam scores are normalized with mean zero and standard deviation equal to one.

Table 2.4: Endogenous Peer Effects on Post-secondary Decisions (studymates)

	SAR (1)	SAR (FE) (2)	SCSAR (FE, $\bar{d}=1$) (3)	SCSAR (FE, $\bar{d}=2$) (4)
Post-secondary Aspirations				
Enrollment	0.033	0.015	0.001	-0.006
	[0.03, 0.05]	[0.00, 0.02]	[-0.01, 0.01]	[-0.02, 0.01]
Private college	0.054	0.024	0.024	0.015
	[0.04, 0.07]	[0.01, 0.04]	[0.01, 0.04]	[0.00, 0.03]
Top-10 college	0.066	0.034	0.032	0.029
	[0.05, 0.08]	[0.02, 0.05]	[0.02, 0.04]	[0.02, 0.04]
Academic (4 year) degree	0.050	0.020	0.017	0.011
	[0.04, 0.06]	[0.01, 0.03]	[0.00, 0.03]	[-0.01, 0.03]
STEM degree	0.043	0.022	0.021	-0.014
	[0.03, 0.06]	[0.01, 0.04]	[0.01, 0.03]	[-0.03, 0.00]
Expected wage	0.035	0.011	0.002	-0.002
	[0.02, 0.05]	[-0.00, 0.03]	[-0.01, 0.02]	[-0.02, 0.01]
Enrollment Choices				
Enrollment	0.053	0.006	0.003	-0.011
	[0.04, 0.07]	[-0.01, 0.02]	[-0.01, 0.02]	[-0.03, 0.00]
Private college	0.022	-0.004	-0.005	-0.025
	[0.01, 0.03]	[-0.02, 0.01]	[-0.02, 0.01]	[-0.05, -0.00]
Top-10 college	0.024	-0.001	-0.001	-0.008
	[0.01, 0.04]	[-0.01, 0.01]	[-0.02, 0.01]	[-0.02, -0.00]
Academic (4 year) degree	0.051	0.020	0.020	-0.004
	[0.04, 0.06]	[0.01, 0.03]	[0.01, 0.03]	[-0.03, 0.01]
STEM degree	0.039	0.013	0.012	0.006
	[0.03, 0.05]	[0.00, 0.03]	[-0.00, 0.03]	[-0.01, 0.02]
Expected wage	0.054	0.015	0.013	0.001
	[0.04, 0.07]	[0.00, 0.03]	[-0.00, 0.03]	[-0.01, 0.02]
Exit Exam				
Overall score	0.083	0.045	0.036	0.035
	[0.07, 0.09]	[0.03, 0.06]	[0.02, 0.05]	[0.02, 0.05]
Mathematics	0.051	0.010	0.004	0.000
	[0.04, 0.06]	[-0.00, 0.02]	[-0.01, 0.02]	[-0.02, 0.01]
Language	0.059	0.030	0.021	0.014
	[0.05, 0.07]	[0.02, 0.04]	[0.01, 0.03]	[-0.00, 0.03]
Sciences	0.072	0.036	0.026	0.020
	[0.06, 0.08]	[0.02, 0.05]	[0.01, 0.04]	[0.01, 0.03]
Social sciences	0.067	0.028	0.019	0.019
	[0.05, 0.08]	[0.02, 0.04]	[0.01, 0.03]	[0.00, 0.03]

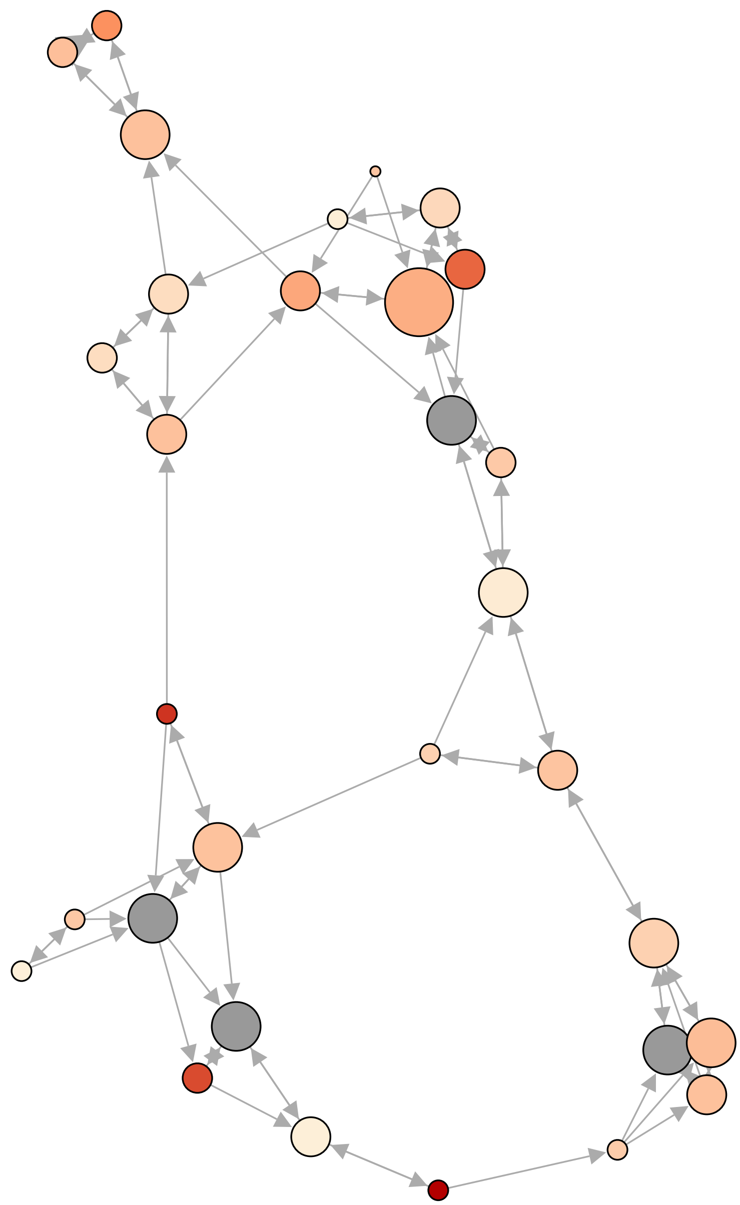
Note: Results are based on a sub-sample of 194 networks and 5,172 students that include all students with at least one study mate in the sample. Each cell represents a separate Bayesian estimations based on 20,000 replications with 5,000 burn-in steps. Posterior means and 95% confidence intervals (in brackets) of the endogenous peer effects are reported. Expected wages are expressed in minimum wages and exit exam scores are normalized with mean zero and standard deviation equal to one.

Table 2.5: Endogenous Peer Effects on Post-secondary Decisions (Reciprocal nominations)

	SAR (1)	SAR (FE) (2)	SCSAR (FE, $\bar{d} = 1$) (3)	SCSAR (FE, $\bar{d} = 2$) (4)
Post-secondary Aspirations				
Enrollment	0.037	0.010	0.006	0.006
	[0.03, 0.05]	[0.00, 0.02]	[-0.01, 0.02]	[-0.01, 0.02]
Private college	0.051	0.022	0.022	-0.002
	[0.04, 0.06]	[0.01, 0.04]	[0.01, 0.04]	[-0.02, 0.01]
Top-10 college	0.067	0.037	0.036	0.025
	[0.05, 0.08]	[0.02, 0.05]	[0.02, 0.05]	[0.01, 0.04]
Academic (4 year) degree	0.052	0.026	0.023	0.020
	[0.04, 0.07]	[0.01, 0.04]	[0.01, 0.04]	[0.01, 0.03]
STEM degree	0.040	0.022	0.020	0.023
	[0.03, 0.05]	[0.01, 0.04]	[0.01, 0.03]	[0.01, 0.04]
Expected wage	0.036	0.014	0.010	-0.005
	[0.02, 0.05]	[0.00, 0.03]	[-0.00, 0.02]	[-0.02, 0.01]
Enrollment Choices				
Enrollment	0.045	0.003	-0.005	-0.011
	[0.03, 0.06]	[-0.01, 0.02]	[-0.02, 0.01]	[-0.03, 0.00]
Private college	0.021	-0.004	-0.019	-0.002
	[0.01, 0.03]	[-0.02, 0.01]	[-0.03, -0.01]	[-0.02, 0.01]
Top-10 college	0.025	0.006	0.003	-0.008
	[0.01, 0.04]	[-0.01, 0.02]	[-0.01, 0.02]	[-0.03, 0.01]
Academic (4 year) degree	0.048	0.021	0.011	0.019
	[0.04, 0.06]	[0.01, 0.03]	[-0.00, 0.02]	[0.01, 0.03]
STEM degree	0.038	0.015	0.009	-0.006
	[0.02, 0.05]	[0.00, 0.03]	[-0.01, 0.02]	[-0.02, 0.01]
Expected wage	0.049	0.017	-0.001	0.012
	[0.04, 0.06]	[0.00, 0.03]	[-0.02, 0.01]	[-0.00, 0.03]
Exit Exam				
Overall score	0.077	0.042	0.037	0.027
	[0.06, 0.09]	[0.03, 0.06]	[0.02, 0.05]	[0.01, 0.04]
Mathematics	0.042	0.006	0.003	-0.033
	[0.03, 0.05]	[-0.01, 0.02]	[-0.01, 0.02]	[-0.05, -0.02]
Language	0.053	0.027	0.021	0.005
	[0.04, 0.07]	[0.01, 0.04]	[0.01, 0.03]	[-0.01, 0.02]
Sciences	0.068	0.036	0.031	0.000
	[0.06, 0.08]	[0.02, 0.05]	[0.02, 0.04]	[-0.02, 0.01]
Social sciences	0.060	0.025	0.020	0.015
	[0.05, 0.07]	[0.01, 0.04]	[0.01, 0.03]	[0.00, 0.03]

Note: Results are based on a sub-sample of 194 networks and 5,109 students that include all students with at least one reciprocal nomination in the sample. Each cell represents a separate Bayesian estimations based on 20,000 replications with 5,000 burn-in steps. Posterior means and 95% confidence intervals (in brackets) of the endogenous peer effects are reported. Expected wages are expressed in minimum wages and exit exam scores are normalized with mean zero and standard deviation equal to one.

Figure 2.1: A Classroom Network Example



Notes: Network with 31 students (nodes) and 83 nominations (directed edges). The color of nodes represents aspired wages (warmer colors for higher wages, and grey for missing data) and the diameter corresponds to in-degree, i.e. number of nominations a student receives. Network visualization is done with Gephi (Bastian et al., 2009).

Chapter 3

Local Effects of Small-Scale Mining on School Education and Child Labor: Evidence from the Colombia's Gold Rush

3.1 Introduction

According to official statistics, Colombia's gold production tripled between 2001 and 2012. The area covered by mining titles grew at even faster rates. This boom was mostly driven by a sharp rise in international prices, set off by the 2007-2009 financial crisis. However, unlike other large producers, most of this gold in Colombia was produced by small-scale artisan and illegal miners. In fact, over 87% of the gold mines reported having no title in the 2010-2011 Mining Census. What should have been a windfall has proven to be in many cases harmful to local development. For instance, deforestation and health hazards related to mercury contamination of water sources have dramatically increased in mining areas (e.g. IDEAM, 2015, Cordy et al., 2011, Romero and Saavedra, 2015). Illegal mining has also become a major source of financing for illegal armed groups, which has intensified the internal conflict (e.g. Massé and Camargo, 2012, Dube and Vargas, 2013, Idrobo et al., 2013, Rettberg and Ortiz-Riomalo, 2014).

Abundant qualitative evidence indicates that dropout rates and child labor have also increased in gold mining areas (e.g. Defensoría del Pueblo, 2010, Gonzalez et al., 2013, Goñi et al., 2014, El Tiempo, 2013, El Espectador, 2013b). This is consistent with the education channel of the "natural resource curse" hypothesis: mining can raise the opportunity cost of studying, and affect long-run economic development by reducing the accumulation of human capital. This may happen if children work on mining. In Colombia, case studies show that both boys and girls participate in gold mining and related activities, and that the probabilities of joining the labor force are higher after age 10. While most of the children combine school and work, those who work are more likely to drop out (e.g. ILO et al., 2001b, ICBF, 2001). Consistently, the International Labor Organization (ILO) has classified small-scale mining as one of the worst forms of child labor in Colombia, and estimates that it employed over 200,000 children in 2000 ILO et al. (2001a).[1] It could also be the case that parents in mining areas are working more, and therefore, children are doing additional household labor.

[1] According to the Colombian Law, school attendance is compulsory up to age 15 or grade 9, and children are allowed to work after 15 with parents' consent. Although there has been some progress over the last decades, the enforcement of the law is still limited, especially in the poorest regions. The 2005 Population Census indicates that 19% of the children aged 5-17 do not go to school, with particularly low attendance rates for the groups 5-6 (71%) and 16-17 (61%). The 2011 Child Labor Survey indicates that 1,465,031 children are working, representing 15.4% of the population aged 5 to 17. The share of workers is much higher for males (16.6%), rural areas (19.1%) and the age group 15-17 (27.7%).

This paper provides new empirical evidence of the impacts of mining on school education and child labor, in a context where small-scale artisan and illegal mining are predominant. I propose two measures of annual change in mining intensity in the proximity of each school and household. The first one is the area covered by active mining titles, and captures the expansion of legal mining. The second one is the deforestation in areas with identified gold deposits, which is a proxy for all mining activities, whether they are legal or not. To correct for measurement error and potential endogeneity problems, I instrument the mining intensity measures with the interaction between gold deposits in the area and international prices. Education and child labor outcomes are observed from both the school and household perspective by combining three different sources of information: School administrative records, national exit exams, and Demographic and Health Surveys (DHS). The analysis is based on municipalities under 200.000 inhabitants located in Antioquia and the Coffee Region departments (Caldas, Quindío and Risaralda). This region accounts for over 55% of the reported gold production and satisfies two conditions that are critical for the empirical strategy: 1. The schools are geocoded; 2. Gold deposits are well identified, and there is enough variation in local mining intensity.

The main results indicate that mining significantly increases dropout rates in urban areas, with larger effects for females in primary. An additional standard deviation in mining increases dropouts rates up to 10.1 percentage points. There are also some positive impacts on repetition rates for males. For younger children, this is partially driven by higher labor participation. In fact, the DHS surveys reveal that mining reduces school attendance, and increases the probability of working of children aged 9 to 11, with estimates effects as large as 9.4 percentage points. The effect on working is also positive for older age groups, although not statistically significant. Effects on exit exams performance are limited. Interestingly, the active titles measure yields smaller, yet significant, estimates for most outcomes. This has two implications. First, in this context even legal mining had perverse effects on children. This cast doubts on the capacity of local authorities to alleviate the negative externalities of mining. In particular, it raises concern on the effectiveness of the royalties system, which allocates additional resources to mining municipalities to be invested primarily on education and health. Second, the effects are larger when artisan and illegal mining are accounted for. Efforts towards formalizing artisan miners, and controlling illegal mining are therefore expected reduce dropouts and child labor.

The paper contributes to the literature in at least two ways. First, the growing research on the local economic effects of mining has mostly focused on large-scale projects (e.g. Aragón and Rud, 2013, Wilson, 2012, Kotsadam and Tolonen, 2015, Chuhan-Pole et al., 2015). This is on of the few papers that study the local effects of mining in a context where small-scale artisan and illegal miners are predominant. Recall that small-scale gold mining employs over 20 million workers in the world, and accounts for approximately 15% of the total production (UNEP, 2015). The proposed measures of local mining intensity are adapted to the nature of small-scale mining and can be replicated in other countries. Second, while most of the empirical evidence on the effect of aggregate economic shocks on human capital accumulation is based on country-level analysis (e.g. Sachs and Warner, 1995, Gylfason, 2001, Stijns, 2006), there are relatively few papers studying this problem at a sub-national level. The most closely related paper is Santos (2014), who finds that gold mining reduced school attendance and increased child labor in Colombia between 1993 and 2005. The author uses the IPUMS samples from the 1985, 1993 and 2005 population censuses, and measures mining at the municipal level with the interaction between gold capability and international prices. This paper finds similar results for a more recent period, 2004-2012, during which the country witnessed the biggest gold rush in its recent history. Besides, the paper proposes new measures of local mining intensity, that are more precise and allow studying the differential effects of legal and illegal mining.

The remainder of this chapter is organized as follows. The next section briefly introduces the Colombian gold mining sector, emphasizing the prevalence of small-scale artisan and illegal mining. Section 4 describes the data and empirical strategy. Section 5 presents the main results, and the last section concludes.

3.2 Prevalence of Small-Scale Mining in Colombia

Colombia experienced a gold rush over the past ten years. According to official statistics, gold production grew from 21 tons in 2001, to over 65 tons in 2012 (Panel (a) of Figure 3.1). In 2012, the country was the 11th largest producer of gold in the world, surpassing Brazil and Indonesia (USGS, 2013). The boom, that was not exclusive to Colombia, was driven by a sharp rise in international prices, mostly caused by the financial crisis; For the first time since the eighties, the priced of gold passed 600 USD per troy ounce, reaching a maximum of 1,900 USD in September 2011. As it has been extensively documented, during this period gold acted as a safe haven for financial assets, which dramatically increased demand and prices (e.g. Baur and McDermott, 2010, Reboredo, 2013). Colombia also witnessed an accelerated growth of mining titles.[2] The area covered by approved gold titles climbed from 3,583 km^2 in 2001, to 27,290 km^2 in 2012 (Panel (b) of Figure 3.1). Although this boom was mostly motivated by high prices, there were also generous tax incentives and legislative reforms designed to attract foreign investors, who now own most of the titles.[3] The title expedition process, however, was far from transparent. Corruption scandals and lack of administrative capacity forced the government to stop the application process between 2011 and 2013, and restructure the title expedition process and the Mining Cadastre.[4]

In spite of the rapid growth of mining titles, Colombia's gold sector is still dominated by small-scale mining, and in most of the activity fails to fulfill legal requirements. The most recent Mining Census, that took place in 2010-2011, reveals that 87% of the gold mines operate without a title, and only 3% have mandatory environmental permits. The reality

[2] Colombian subsoil resources are property of the state and mining firms are granted titles to exploit them. Since 2001 (Mining Code, Law 685 of 2001), Concession Agreements are the only legal form of contracting. These titles, granted for up to 30 years, contemplate three phases: exploration, construction and exploitation. The exploration phase period is 3 years, and can be extended up to 11 years. Environmental permits are required to begin the construction and exploitation phases. During the initial phases, firms pay a yearly fee that is determined by the surface area. Once the exploitation phase begins, mining companies pay royalties over the reported production (4% for gold and silver, and 6% for alluvial gold). Royalties are collected by the central government, and mining departments and municipalities receive a fraction of them to be invested primarily on education and health projects.

[3] Between 2002 and 2005, the Congress approved sector-specific income tax breaks, tax deductions for investments, and legal stability contracts. Also, the new Mining Code simplified the institutional framework for doing business.

[4] The corruption scandals included several cases of bribery, expedition of titles in restricted areas, and speculation with mining titles. By 2011, there were 19,000 accumulated requests, most of which were eventually rejected by the new National Mining Agency (The Economist, 2013).

is probably worse; compliance to the Census was not mandatory, and officers failed to visit mines in high-conflict regions. An unknown share of the untitled production is sold to local traders and exporting companies, which pay the corresponding royalties and export it. Goñi et al. (2014) estimate that 23% of the gold is sold in a different municipality. In this process, all possibility of tracking the origin of the gold is lost. These statistics reflect two failures of the Colombian mining policy. One the one hand, the various plans to formalize labor-intensive artisan mines have systematically failed (Echavarria, 2014, Gonzalez et al., 2013, Defensoría del Pueblo, 2010). On the other hand, illegal alluvial mining, which is highly mechanized, has rapidly expanded throughout the map. While artisan miners have existed for centuries and have never been a priority of the government, illegal mining has become a major concern. Over the last few year, the government has significantly increased the number of raids, and introduced legislative reforms designed to facilitate the enforcement of the law.[5]

The recent mining boom has had devastating effects on the environment. Alvarez-Berríos and Aide (2015) estimate that the impact of mining on deforestation in South America significantly increased over the past years, reaching 116,000 Ha of lost forest in gold mining areas between 2007 and 2013. The authors identify a high-deforestation cluster in the mining area located in the northeast of Antioquia. Consistently, official statistics indicate that mining is nowadays the primary cause of deforestation in Antioquia and Chocó

[5] For instance, the Law 1453 of 2011 increases the penalty for illegal mining and environmental damage. Likewise, the Decision 774 of 2012, by the Andean Community (CAN) (and the regulatory decrees) restrict the machinery commerce, create a unified mining register and allow the authorities to seize or destroy machinery in absence of mining titles.

(IDEAM, 2015). Abundant evidence has also shown that mercury contamination of water sources has increased, and so have the number of reported cases of mercury poisoning (e.g. IDEAM, 2015, Cordy et al., 2011, Güiza and Aristizabal, 2013). Romero and Saavedra (2015) show that mercury has also affected the health of newborns.

Gold mining has also become a security threat. Mining is a growing source of financing for illegal armed groups, which has intensified the conflict in mining regions (e.g. Massé and Camargo, 2012, Dube and Vargas, 2013, Idrobo et al., 2013, Rettberg and Ortiz-Riomalo, 2014).[6] Official reports also indicate that forced displacement and human rights violations have increased in municipalities with large-scale mining projects (e.g. Garay et al., 2013). Besides, illegal mining is a constant source of tax fraud. In fact, miners and traders systematically evade royalties and income taxes, which seriously affects the revenue of local authorities (e.g. Garay et al., 2013, Portafolio, 2011, El Colombiano, 2012). There is also evidence of royalties fraud schemes, where mafias falsely report gold as produced in municipalities where they have control on local authorities and public spending(El Espectador, 2013a). Moreover, ongoing investigations indicate that gold has been used as part of a massive money laundering schemes (The Telegraph, 2016, Bloomberg, 2015).

Antioquia is the epicenter of gold mining in Colombia. The department accounts for 52% of the reported 2001-2012 production, and 33% of the titled area. Including the Coffee Region departments, the share of production and titled area is 55% and 37%. There are

[6] Anecdotal evidence and police reports indicate that most miners and traders pay extortions in conflict areas, and that illegal armed groups are in many cases involved in mining activities. The strong geographical correlation between illegal mining and coca plantations documented by UNODC (2015) partially confirms this.

four mining clusters in this region (See the requested and approved mining titles in Figure 3.2). The first and most important one is in the northeast of Antioquia, and south of Bolívar. The second one is located in the southwest of Antioquia, and Caldas, Risaralda and Chocó. The third and fourth are in the West of Antioquia, and the East of Quindío and Tolima. It is worth noting that there has been gold mining in these four areas since the colonial period, and there is evidence indicating that this has negatively affected long term economic development and human capital accumulation Acemoglu et al. (2012). However, mining activity in the West of Antioquia and Tolima (clusters 3 and 4) declined throughout the twentieth century, leaving only some artisan and small-scale operation behind. The numerous titles in these two areas correspond to ongoing large-scale projects that have not reached the production stage.[7] The region combines underground mining in the mountains and alluvial mining in the river beds. While underground mining requires large investments and has grown at a relatively slow pace, alluvial mining rapidly expanded during the period of study.

3.3 Data and Methods

This paper estimates the effect of gold mining on a set of education and child labor outcomes. I focus on Antioquia and the Coffee Region departments -Caldas, Quindío and Risaralda, because they satisfy two conditions that are critical to the empirical strategy.

[7] In particular, there are two open-pit projects, *Mandó Norte* and *La Colosa*, that are in exploratory and feasibility phase. These projects have been largely criticized for human rights violations, irregularities in the prior consultation process, and environmental impacts. Multiple legal actions, including local referendums, have been taken against the projects (e.g. Campaign, 2013, Cárdenas, 2014).

First, the available information allows geocoding most of the schools. second, gold deposits are well identified and there is enough variability in the intensity of mining. Given that mining represents only a small share of the economy of large cities, I further restrict the sample to municipalities smaller than 200,000 that are not part of a metropolitan area. This section first presents the data, emphasizing each of the conditions above, and then describes the proposed mining intensity measures and the empirical strategy.

3.3.1 Education and Child Labor

Education and child labor are measured using three different sources of information that provide complementary evidence of the effects of gold mining. School administrative records follow enrollment, grade promotion, repetition, and dropout rates at different education levels. Exit exam databases have individual information on test scores and working situation of senior high school students. DHS surveys, focus on households and allow estimating the effect of mining on school attendance and child labor.

School administrative records are collected annually by the National Statistic Department (DANE). School principals are required to complete the *C600* form, which includes questions on previous year enrollment, grade promotion and repetition, dropouts, and transfers since 2005. Using the 2005-2013 datasets, I measure the initial enrollment and progress throughout the year by level of education for the period 2004-2012. Each school may have multiple shifts, in which case they are treated separately. Excluding large cities and metropolitan areas, there are 56,116 schools, of which 36,713 have information for at least eight years. These schools are for the most public (94.6%) and rural (78.8%). Primary schools are not always separate from middle schools and high schools. 95% of the schools in the sample offer primary education (grades 1-6), of which 20% also have middle school (7-9) and high school (10-11). Descriptive statistics of initial enrollment and progress throughout the year are presented in Panel A of Table 3.1. Primary schools enroll on average 73 students at the beginning of each school year, of which 82.7% get promoted, 6.7% repeat, 6.9% dropout, and the remaining 3.7% transfer to other schools. There are much less middle and high schools, but each one of them enrolls more students. Dropout rates are higher in middle school (7.4%), than in high school (4.7%), which reflects that students who enroll in high school are a relatively self-selected group. Recall that schooling is mandatory until age 15 or grade 9.

National exit exams -*SABER 11*- are administered by the Colombian Institute for the Promotion of Higher Education (ICFES). They are required for college application and in some cases for graduation as well. Most senior students take them. Anonymized individual test scores are available since 2000 but I restrict the sample to the 2004-2012 period to have comparable results. While registering for the exam, students complete a survey that provides basic information such as date of birth and gender.[8] Since 2008, students are also asked about their labor situation. Excluding metropolitan areas, there are 6,245 schools, of which 2,703 have at students taking the exam in at least eight years. The share of public and rural schools is smaller at this level, with 83.8% and 40%, respectively. As can be seen

[8] Other characteristics, such as parents' education and family income are not included in some of the years of analysis, and therefore cannot be used as controls.

in Table 3.1, schools have on average 52.2 students taking the exam every year. The test scores are slightly bellow the national mean 0.013 standard deviations) and 11% of the students report working.

The Ministry of Education, in coordination with the Secretaries of Education of each department, has geocoded over 24,000 schools in the country. There are nine departments in which more than 85% of the schools (excluding metropolitan municipalities) are geocoded: four in the central region -Antioquia, Caldas, Quindío and Risaralda- and five in the Orinoco-Amazon region -Meta, Putumayo, Guanía, Guaviare and vaupés. I focus on the four departments from the central region for two main reasons. First, they are historically connected and have comparable living standards. Caldas, Quindío and Risaralda were in fact colonized by Antioquian settlers during the nineteenth century, and they all benefited from the Coffee booms throughout the twentieths century. According to the 2005 Census, the poverty rates of the four departments oscillate between 16.2 and 22.9%, which is considerably bellow the Orinoco-Amazon region (39.9%) and the national average (27,7%). Second, the region has some of the biggest mining clusters in the country, but also enough municipalities without this activity. This is not the case of the Orinoco-Amazon region where there are relatively few mines. This point is further developed in the next subsection.

Antioquia and the Coffee Region account for 17% of the schools and 19% of the students taking the exit exam. The region has higher repetition and transfer rates than the national sample. Senior students also have slightly lower test scores, and higher probabilities of working (13%). (See Table 3.1). Approximately half of the schools without coordinates are classified as urban in the *C600* records. In these cases, I impute the coordinates of the corresponding municipal towns. With this correction, over 97% of the schools in the region of study are geocoded. While most of the schools are located in the the central region of Antioquia, and the Coffee Region, there are relatively few schools in the Northeast and West of Antioquia, which reflects the low population density of these areas (Panel (a) of Figure 3.3).

The 2005 and 2010 waves of the DHS provide information on school attendance and labor situation of a sample of children aged 6 to 17. While the 2010 wave has GPS information, in 2005 it is only possible to accurately geocode the clusters classified as urban. I do so by imputing the coordinates of the corresponding municipal towns. Given the data limitations, the main estimates consider only households living in the urban area of municipalities under 200.000. There are 4,261 children aged 6 to 17 in non-metropolitan municipalities of Antioquia and the Coffee Region, representing 11% of the total sample. On average, 84% of the children go to school. There is some late entrance, only 66% of children under 8 attend school, and dropouts increase after 15. The percentage of children working increases with age; less than 3% work under 11, 8% in the group 12-14, and 19% over 15. The location of all the geocoded DHS clusters are presented in Panel (a) of Figure 3.3. Although there are less clusters, their spatial distribution is similar to that of the schools (Panel (a) of Figure 3.3).

3.3.2 Measuring Mining Intensity

One of the practical consequences of the prevalence of small-scale artisan and illegal mining is that the existing measures of gold production are limited and unreliable. For instance, official production statistics, based on royalties, not only fail to account for an unknown fraction of the illegal mining, but are also distorted by royalties frauds and money laundering schemes (see Section 3.2). Besides, production is aggregated at the municipal level, and there is no way to track it to a particular mine. The Mining Cadastre, which registers all approved and requested titles, also has several limitations. Notably, there is no information about the production of each mine. Moreover, approval dates are not always good predictors of gold production. In fact, there are some areas where artisan and illegal miners have operated long before the license was requested. There are also mines with approved concession contracts that have not reached the production phase.

This paper proposes two new measures that capture annual changes in mining intensity in the proximity of each school and household. The first one is the area covered by *active* titles. A title is considered active if it was approved during the year (or before) and has not expired. This measure is intended to capture the expansion of legal mining, independently of the stage of the project. The second measure is the annual deforestation in areas with identified gold deposits. This is a proxy for all mining activities, whether they are legal or not. These measures are calculated using detailed geographic information on mining titles and deforestation.

The mining title information is obtained from three different sources. First, the Mining Information System of Colombia (SIMCO) which has administrative information of all mining titles.[9] Second, the Colombian Geographic Information System for Planning (SIGOT) which provides geographic information of all titles approved until 2012.[10] Third, *TierraMinada*, an NGO that collected administrative and geographic information on both approved and requested mining titles up to November 2014. [11] When titles cover more than one municipality, they are divided and treated separately. There are in total 3,229 approved titles for gold, and 839 pending requests. 54% of the approved titles, and 41% of the requests are in Antioquia or the Coffee Region (See Figure 3.2).

Deforestation statistics are calculated using the Hansen et al. (2013) high-resolution forests cover maps. The authors estimate the annual forest loss at spatial resolution of 30 meters for the period 2001-2012.[12] Between 2001 and 2012, the country lost nearly 3 million Ha of forest. The expansion of the Amazon agricultural frontiers accounts for more than half of it. The second largest deforestation hotspot is located in the Northeast and East of Antioquia (See Figure B.1 of the Appendix). As mentioned in section 3.2, mining has become the primary source of deforestation in this area (IDEAM, 2015).

The areas with gold deposits are delimited using all the *approved* and *requested* mining titles up to 2014. There are three arguments to justify this choice. First, the title expedition

[9] The data can be downloaded from `http://www.simco.gov.co/Inicio/CatastroMineroColombiano/tabid/107/Default.aspx`.

[10] The shapefiles can be downloaded from `http://sigotn.igac.gov.co/`.

[11] The shapefiles and datasets can be downloaded from `https://sites.google.com/site/tierraminada/`.

[12] Imagery for 2013 onward is also available, however forest loss measures are not comparable due to methodological changes. The maps can be downloaded from `https://earthenginepartners.appspot.com/science-2013-global-forest`.

process is expensive, and there are no incentives for investing in areas with no mining potential. Consistently, during the period of study, titles were often requested in areas traditionally exploited by artisan miners (Goñi et al., 2014, Gonzalez et al., 2013). Second, anecdotal evidence suggests that there is abundant illegal mining in areas that have pending title requests. This is particularly true in region of study, where the UNODC (2015) remote sensing evidence of alluvial mining consistently overlap with mining titles. Third, compared to the Pacific and Orinoco-Amazon regions, there are relatively few protected areas in Antioquia and the Coffee Region, that may prevent or delay the title expedition (See Figure B.1 in the Appendix).[13]

I calculate the distance between each mining title and all the schools and DHS clusters in the sample. The main set of regressions assume that a mine is in the neighborhood of a school (or household) if the distance is inferior to 20 km. In this context, miners are not expected to commute longer distances on a daily basis. I also test the sensibility of the results to alternative distance buffers ranging from 10 to 50 km. The active titles measure is the total area covered by titles in the neighborhood of a school (or household) s that are active in year t. Figure 3.4 shows the distribution of this measure. While 20-25 % of the sample has no active licenses nearby, the schools and households are surrounded on average by 7.9 and 9.1 km^2 of active titles, respectively.

[13] Mining activity is strictly forbidden in National Parks. Although some titles were requested and approved in these areas over the last few years, these remain exceptional cases. In Indigenous Reserves and Afro-descendant Territories, mining projects are subject to Prior Consultation. In practice, the enjoyment of this right remains limited: there is evidence of systematic violence against ethnic groups in mining areas, and multiple projects have been implemented without fulfilling this requirement (DPLF, 2015).

The area covered by approved and requested mining titles in the neighborhood of each school (or household), dep_s, is a time-invariant measure that captures the presence of gold deposits. Using the deforestation and mining titles maps, I calculate the annual forest loss in each titled area. Antioquia and the Coffee Region departments account for 58% of the total deforestation in mining areas between 2001 and 2012, with a particularly high concentration in the Northeast (See Panel (b) of Figure 3.3). The second local measure of mining intensity, hereafter referred as mining deforestation, is the deforestation in mining areas in the neighborhood of the school (or household) s in yeat t. The average annual mining deforestation in the 20 km of schools and households is 9 and 8.2 Ha, respectively. Compared to active titles, there are more schools and households with positive mining deforestation (See Figure 3.5). In both cases, there is enough variability in the mining intensity measures in the region of study.

3.3.3 Emprical Strategy

The aim of this paper is to measure the local effect of gold mining on education and child labor. I use a difference-in-differences approach that exploits the spatial and temporal variation of the mining intensity measures. For simplicity of interpretation, the mining intensity measures are normalized with mean zero and standard deviation one in each sample. The estimated coefficients should therefore be interpreted as the effect of one standard deviation increase in mining intensity. The specification of the model depends on the unit of analysis: school outcomes (e.g. enrollment, dropout rates, and number of students taking the exit exam) are estimated with panel fixed-effects models, and individual outcomes (e.g. test scores, school attendance, and child labor) use repeated cross-section methods.

The panel fixed-effects model, presented in Equation 3.1, regresses the outcomes y_{smt} of school s, municipality m and year t, on the gold mining measure of choice $Gold_{st}$. School fixed effect (μ_s) capture observed and unobserved school and location characteristics that may affect the outcome, including the gold deposits in the area. Year fixed effects (τ_t) and linear municipal-specific time trends ($\eta * t$) control for common shocks, including trends in the municipal fiscal revenue, public spending and conflict. Errors are clustered at the school level.

$$y_{smt} = \beta_0 + \gamma Gold_{st} + \mu_s + \tau_t + \eta_s * t + \epsilon_{smt} \tag{3.1}$$

The specification used for exit exam individual outcomes is relatively similar. The regressions control for school and year fixed effects and municipal-specific time trends. I also control for the students' age and gender. Errors are clustered at the school level (Equation 3.2).

$$y_{ismt} = \beta_0 + \beta_1 X_{ismt} + \gamma Gold_{st} + \mu_s + \tau_t + \eta_s * t + \epsilon_{ismt} \tag{3.2}$$

The last set of regressions are based on the 2005 and 2010 DHS waves. I match for each individual i in cluster s the mining intensity measures of the corresponding year. The regressions control for municipal and year fixed effects, and a set of individual characteristics including age, gender, household size and parents' education.[14] They also control for gold deposits in the neighborhood of each cluster, dep_s, to capture the time-invariant effect of living in areas with high potential for gold mining. I use the DHS sample weights, and errors are clustered at the DHS cluster level (Equation 3.3).

$$y_{ismt} = \beta_0 + \beta_1 X_{ismt} + \gamma Gold_{st} + \lambda dep_s + \mu_m + \tau_t + \epsilon_{ismt} \tag{3.3}$$

There are two potential sources of bias that need to be considered. First, mining intensity is measured with error. In fact, active mines only account for legal mining and do not differentiate between the different stages of the mining projects. Likewise, deforestation in areas with known deposits can be caused by other activities, and it is not possible to perfectly delimit gold deposits. Second, there could be some unobserved factors that determine mining activities and also affect educational and child labor outcomes. For instance, variation in the intensity of the armed conflict not captured by the fixed effects and the municipal-specific time trends. Both of these problems are addressed using an instrumental variable approach.

[14] Parents' education is the higher education level completed by the parents, or the household head if both parents are absent. The education level is categorized in three groups: primary, secondary and higher education.

I instrument the gold mining measures with the interaction between gold deposits in the neighborhood and international prices of gold ($dep_s \times P_t$). Notice that the separate effects of gold deposits and prices are absorbed by the school and year fixed effects (or the variable dep_s in the DHS regressions). The exclusion restriction is satisfied by the exogeneity of gold deposits and international prices. Gold deposits are determined by geological and geochemical properties of the land. As discussed previously, there are good reasons to believe that they are well delimited by mining titles (requested or approved) in the region of study. Prices are determined in international markets. Even though Colombia significantly increased its production during the period of study, the country is still considered a price taker. Besides, most of the variation in international prices during the period of study was driven by the financial crisis. The results presented in the next section confirm that this instrument is strongly correlated with both mining intensity measures.

3.4 Results

The main results of the paper are presented in three parts. First I study the impacts of mining from the school perspective using school administrative records and exit exams. The second part is based on DHS surveys and focuses on the children and their families, whether they are studying or not. Although presented separately, the evidence from these sources should be interpreted as complementary. The third part tests the robustness of the results to different distance buffers, and also to controlling for two potential confounding factors: homicides and royalties.

3.4.1 School Administrative Records and Exit Exams

The effects of mining on school enrollment and progress throughout the year are presented in Table 3.2. The regressions are estimated separately for primary, middle school and high school, and the mining intensity measures are based on 20 km neighborhoods. The first thing to be noticed is that while most of the OLS coefficients are not significant, the IV regressions yield estimates that are larger in magnitude and significance. Besides, the instrument is highly correlated with both of the mining intensity measures; the first-stage F-statistics oscillate between 37.9 and 600.8 and are statistically significant. The first-stage estimates, presented in Table B.1 of the Appendix, show that the instrument has positive and significant effects on both of the mining intensity measures (0.52 to 0.59 for active titles, and 0.25 to 0.31 for mining deforestation). This confirms the importance of correcting for measurement error and potential endogeneity.

The main IV estimates indicate that gold mining does not affect the enrollment at the beginning of each year, however, it reduces the number of students promoted, and increases the dropout and repetition rates in primary and high school. These effects are quite large; an additional standard deviation in active titles reduces the primary and high school promotion rates by 2.3 and 4.7 percentage point, respectively. The effects are even larger for mining deforestation, with estimated effects of 5.4 and 7.8 percentage points. Consistently, mining increases primary dropout rates by 1.3 to 2.9 percentage points, and high school dropout and repetition rates by 2.9 to 5.3 percentage points.

Interestingly, the increase in dropouts is not followed by lower enrollment. I estimate the effect on the enrollment of the following year, finding no significant effects (Table B.2 of the Appendix). This implies that there are enough students entering the system each year to compensate the loss. Migrations might partially explain this. In fact, most of the local authorities report increasing migrations into the mining regions (e.g. Goñi et al., 2014). It could also be the case that students who drop out, go back to school after a year or two. Unfortunately, it is not possible to test these hypotheses with the available information.

Most of the negative consequences of mining are concentrated in urban schools. As can be seen in Table 3.3, the effects of mining are much larger for urban schools, and the differences with respect to rural schools are statistically significant. In urban schools, the primary dropout rate increase by 2.9 percentage points for active titles, and 10.2 percentage points for deforestation mining. The effects on middle school promotion and dropout rates are now significant. These results indicate that children living in rural ares are less vulnerable to mining shocks. There are also some differences by gender. As can be seen in Table 3.4, the increase in repetition rate is concentrated on males, with estimated effects oscillating between 3.7 and 7.6 percentage points. As for dropouts, the effects are relatively similar, except for primary where females are more affected.

The effects on exit exam scores are limited. As with school administrative records, the instrument are strongly correlated with the mining intensity measures, with first-stage F-statistics between 37.5 and 308.4 (first-stage estimates are presented in Table B.3 of the Appendix). The number of students increase with mining. The IV estimates indicate that an additional standard deviation of active titles and mining deforestation increases the number of students by 8 and 13.8, respectively (Table 3.5). The additional students are concentrated in rural schools, where the effects can be as large as 17.4 percentage points, equivalent to a 34% change (Table 3.6). This is consistent with the fact that rural high schools don't see dropouts increase, and have positive, although insignificant, effects on enrollment. The positive effect on the number of students is also concentrated on females 3.7). As for the test scores, most of the estimated effects are small and statistically insignificant; the largest coefficient is under 2% of a standard deviation for female students. Similarly, there are no detectable effects on the probability of working. Overall, these results indicate that students finishing high school are less affected by mining, which reflects that this is a relatively self-selected group.

3.4.2 DHS Surveys

The effects of mining on school attendance and child labor are estimated using urban DHS surveys. I estimate the effect on the whole sample and also by age groups. The main results are presented in Table 3.8. Even with two periods, and considerably less observations, the instrument is still strongly correlated with the mining intensity measures. The smallest first-stage F-statistics in the main specification is 51.7, and is statistically different from zero (see Table B.4 of the Appendix for the first-stage estimates). While there are no significant effects on the entire population, children age 9 to 11 are particularly affected. In fact, the IV estimates indicate that an additional standard deviation of active titles and mining deforestation reduced the probability of studying by 4.2 and 7.9 percentage points, respectively. Moreover, mining increases the probability of working in this age group, with estimated effects of 5 and 9.4 percentage points. Considering that only 2.9% of the children are working at this age, these are very large effects.

These finding indicate that the sharp increase of repetition and dropout rates in urban primary schools are partially driven by an earlier entrance to the labor market. For older students, the effect on working is positive but not statistically significant. In order to learn more about the children who are entering the labor market earlier, I estimate the heterogeneous effects of mining by parents' education. I choose this variable because, unlike family income or wealth measures based on physical assets, it is expected to be uncorrelated to a mining shock. The instruments are weak for mining deforestation but not for active titles. The effect of active titles on both school attendance and child labor are only significant for children with educated parents (Table 3.9). One possible interpretation of this result is that a larger fraction of children with uneducated parents are already working, therefore this group is less sensitive to new incentives. I also test for heterogeneous effects by gender in Table 3.10. While the effects on school attendance and child labor are concentrated on males, the difference between genders is not statistically significant.

I test whether these results hold when rural households are also considered. To do this, I geocode the 2005 rural clusters in the corresponding municipal towns. Since the location of the rural households is measured with error in 2005, results need to be interpreted with caution. In estimate the effect of mining with both urban and rural households in Table B.5 of the Appendix. As can be seen, there are two major differences with respect to the urban only results. On the one hand, the effects on studying are no longer significant. This is consistent with the fact that there were no significant changes in dropout rates in rural primary schools. On the other hand, the effects on child labor of the 9-11 group are larger in magnitude, and the overall effect on the probability of working is now significant.

3.4.3 Robutness

I replicate the main IV estimates for neighborhoods defined by distances ranging from 10 to 50 km. Tables 3.11 and 3.12 present the estimated effects on enrollment, promotion, repetition and dropouts rates. The first thing to be noticed is that there are no significant effects for 50 km, and the sensitivity to distance varies depending on the school level. For instance, the effects on promotion and repetition rates for primary school are larger at shorter distances (10-20 km), as opposed to high school, where they reach their peak between 30 and 40 km. This is consistent with the fact that there are less high schools and they are more spatially concentrated. It may also reflect that teenagers are more mobile. As for dropouts, the largest significant effects are reported for 20 km.

The effect on the number of students taking the exit exam is only significant for the 20 km neighborhood at the 10% significance level, which cast doubts on the robustness of the results (Table 3.13). There are also some positive and significant effects on the test score at distances 30 to 40 km. The coefficients oscillate between 0.013 and 0.038 standard deviations, which is a rather small effect. As for working, the estimated effects are close to zero and statistically insignificant in all specifications.

Table 3.14 presents the DHS estimated coefficients for different distance buffers. The effects on studying and working for the 9-11 group are significant for all neighborhood definitions. There are also some negative and significant effects on the overall probability of studying for distances 30 to 40 km. This seems to be driven by the age groups 12-14 and 15-17 who have the largest estimated coefficients, even though they are not statistically significant. Overall, the DHS results are robust to different distance buffers.

Finally, I control for two potential confounding factors. The first one is homicide rate, which reflects changes in the intensity of conflict at the municipal level. The second one is the municipal annual revenue from royalties. Given the additional fiscal revenue, municipalities with abundant legal mining are expected to perform better in terms of education. As can be seen in Tables B.6 to B.8, results are fairly similar when these variables are included. Although these results do not fully rule out that conflict or public expenditure are mediating the effects of mining on schooling and child labor, they suggest that these are not the key mechanisms driving the estimated effects. Besides, the results also indicate that the empirical strategy is controlling for these factors relatively well.

3.5 Conclusions

This paper estimates the local effect of gold mining on schools and child labor in Colombia. I focus on the 2004-2012 period, during which the country witnessed the biggest mining boom in its recent history. One particular aspect of the Colombian gold mining sector is that it is characterized by the prevalence of small-scale artisan and illegal miners. I use detailed geographic information to construct two measures that capture the annual changes in local mining intensity: the area covered by active titles, which captures the evolution of legal mining, and the deforestation in areas with identified deposits, which is a proxy for all mining activities.

The main results indicate that mining increases dropout rates in urban areas at all levels, with larger effects for females in primary. There are also significant effects on repetition rates for males. The effects are not negligible. For instance, an additional standard deviation of active titles and mining deforestation increases the urban primary dropout rate in 2.8, and 10.2 percentage points, respectively. For the younger children, this effect is partially driven by higher probabilities of working. In fact, the DHS surveys reveal a significant increase in the probability of working in the age group 9-11; with estimated effects between 5 to 9.4 percentage points. The children in this age group also reduce the school attendance by 4.2 to 7.9 percentage points. The effect on working is also positive for children aged-12-17, however the coefficients are not statistically significant. In comparison, the effects on the exit exams are limited. The number of students taking the exam increases, but the effect is not robust to different distance buffers. There are also some small improvements in test scores and no detectable effects on the probability of working.

The effects are expected to be different when mining is done legally. This is partially true here, the active titles measure yields smaller estimates than mining deforestation in all cases. In this sense, any efforts towards formalizing the artisan miners and controlling illegal mining should reduce dropouts and child labor. Moreover, the estimated effects of the legal mining measure are significant in most specifications, which implies that in this particular context, even legal mining has been harmful to children. This provides evidence that local authorities have not been able to alleviate the negative externalities of mining on children. In particular, it cast doubts on the effectiveness of the royalties system, which provide additional funds for education and health in mining municipalities. Such findings are consistent with previous literature showing that royalties have, in most cases, failed to improve living standards in mining areas (e.g. Perry and Olivera, 2009, Echeverry et al., 2011, Aguilera et al., 2014, Martınez, 2016).

A number of questions remain unanswered. First, mining dropout rates increase in middle school and high school, even though the effect on working is not statistically significantly for these age groups. The factors behind the dropout decisions in this group are yet to be identified. Second, little is still known about migrations in mining areas. While the qualitative evidence and the enrollment rates suggest that migrants are attracted to mining areas, the magnitude of the flow, and its effects on the composition of local populations are not clear. Third this paper exploits the fact that mining titles allow identifying gold deposits in the region of study. However this is not always the case. For instance, there is evidence of abundant illegal mining in the Pacific region where there less titles have been requested and approved. In order to use this type of measures in such regions, it is necessary to find alternative strategies to precisely delimit areas with gold deposits.

Table 3.1: Descriptive Statistics of the Main Outcomes

		Colombia			Antioquia and Coffee Region		
		Obs.	Mean	(SD)	Obs.	Mean	(SD)
Panel A. School administrative records (2004-2012)							
Enrolled students	Prim.	292,143	73.45	105.42	49,243	75.22	120.12
	Mid.	59,141	204.15	193.79	11,118	187.22	190.13
	High	37,985	97.26	85.46	6,625	96.22	91.77
Promotion rate	Pri.	292,143	82.678	15.355	49,243	78.280	16.485
	Mid.	59,141	83.473	12.664	11,118	81.174	14.749
	High	37,985	88.873	10.064	6,625	87.591	11.160
Repetition rate	Pri.	292,143	6.688	9.769	49,243	9.343	11.988
	Mid.	59,141	6.063	7.297	11,118	6.503	9.272
	High	37,985	4.313	5.758	6,625	4.410	6.794
Dropout rate	Pri.	292,143	6.891	9.626	49,243	6.461	9.119
	Mid.	59,141	7.435	8.769	11,118	7.767	10.169
	High	37,985	4.753	6.651	6,625	4.771	7.575
Transfer rate	Pri.	292,143	3.743	7.544	49,243	5.916	9.076
	Mid.	59,141	3.029	5.873	11,118	4.555	6.992
	High	37,985	2.062	4.272	6,625	3.228	4.710
Panel B. Exit Exam (2004-2012)							
Students per school		23,265	52.042	41.159	5,180	50.830	44.683
Exam score		1,604,203	-0.138	0.902	318,897	-0.197	0.864
Student works*		994,049	0.108	0.310	197,253	0.130	0.336
Panel C. DHS urban households (2005-2010)							
Studies	All	32,748	0.842	0.365	3,627	0.838	0.369
	6-8	7,929	0.664	0.472	846	0.647	0.478
	9-11	8,149	0.965	0.183	868	0.977	0.150
	12-14	8,502	0.949	0.220	963	0.938	0.242
	15-17	8,168	0.780	0.414	950	0.779	0.415
Works	All	32,784	0.075	0.263	3,628	0.080	0.271
	6-8	7,935	0.008	0.089	846	0.008	0.091
	9-11	8,154	0.023	0.151	868	0.029	0.167
	12-14	8,513	0.070	0.255	963	0.082	0.275
	15-17	8,182	0.196	0.397	951	0.187	0.390

Source: Own calculations based on Ministry of Education, DANE, ICFES and DHS.
Note: Cities over 200,000 and metropolitan areas are excluded. Enrollment (Panel A) refers to the number of students registered at the beginning of each school year. The working status of students taking the exit exam are only available only after 2008.

Table 3.2: Effect of Gold Mining on School Enrollment and Progress Throughout the Year (20 km Neighborhood)

	Active titles			Mining deforestation		
	Primary (1)	Middle (2)	High (3)	Primary (4)	Middle (5)	High (6)
Panel A. Enrollment (beginning of the year)						
OLS	-0.228	4.725	5.569**	0.329	-0.071	1.385
	(0.981)	(3.400)	(2.721)	(0.468)	(1.458)	(1.027)
IV	-0.274	4.725	5.841	-0.637	8.876	9.604
	(2.234)	(9.676)	(6.809)	(5.199)	(18.323)	(11.089)
mean(y)	75.911	190.859	98.887	75.911	190.859	98.887
Panel B. Grade promotion rate						
OLS	-0.780*	0.278	0.305	0.306	0.525	0.151
	(0.412)	(0.755)	(0.836)	(0.232)	(0.449)	(0.437)
IV	-2.322**	-1.472	-4.795*	-5.402**	-2.766	-7.884*
	(0.939)	(2.286)	(2.626)	(2.201)	(4.196)	(4.258)
mean(y)	78.249	81.009	87.39	78.249	81.009	87.39
Panel C. Grade repetition rate						
OLS	-0.049	-0.611	-0.234	-0.207	-0.27	0.104
	(0.287)	(0.463)	(0.426)	(0.176)	(0.285)	(0.296)
IV	0.283	0.844	2.913*	0.659	1.585	4.790*
	(0.633)	(1.401)	(1.594)	(1.474)	(2.589)	(2.635)
mean(y)	9.373	6.54	4.499	9.373	6.54	4.499
Panel D. Dropout rate						
OLS	0.264	0.577	-0.02	0.035	-0.276	0.188
	(0.218)	(0.457)	(0.539)	(0.134)	(0.289)	(0.267)
IV	1.244**	1.465	3.207**	2.895**	2.753	5.273**
	(0.509)	(1.550)	(1.612)	(1.192)	(2.872)	(2.564)
mean(y)	6.484	7.886	4.878	6.484	7.886	4.878
First-stage F	600.8	55.5	37.9	666.0	90.6	56.4
Observations	47,785	10,662	6,287	47,740	10,509	6,214

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate panel regression that controls for school and year fixed effects, and municipal-specific time trends. Standard errors are clustered at school level. IV regressions instrument the mining intensity measures with the interaction between gold deposits in the neighborhood and international prices. The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2004-2012. A mine is considered in the neighborhood of a school if the distance is smaller or equal to 20 km. Active titles and mining deforestation, as defined in Section 3.3.2, are normalized with mean zero and standard deviation equal to one.

Table 3.3: Effect of Gold Mining on School Enrollment and Progress Throughout the Year: By Urban/Rural (IV only, 20 km Neighborhood)

	Active titles			Mining deforestation		
	Primary (1)	Middle (2)	High (3)	Primary (4)	Middle (5)	High (6)
Panel A. Enrollment (beginning of the year)						
Rural	0.566	9.838	7.184	1.946	24.975	13.951
	(1.894)	(8.720)	(6.848)	(4.319)	(17.194)	(11.713)
Urban	-5.1	-4.586	3.247	-23.461	-30.578	0.009
	(5.881)	(10.978)	(7.262)	(23.914)	(26.575)	(13.492)
P-value (rural=urban)	0.273	0.024	0.213	0.272	0.018	0.192
Mean(y)	75.911	190.859	98.887	75.911	190.859	98.887
Panel B. Grade promotion rate						
Rural	-1.635*	-0.557	-4.337	-3.267	0.028	-6.607
	(0.948)	(2.335)	(2.726)	(2.216)	(4.520)	(4.554)
Urban	-6.271***	-3.139	-5.680**	-24.263***	-9.612**	-10.703**
	(1.057)	(2.354)	(2.813)	(3.552)	(4.572)	(4.622)
P-value (rural=urban)	0.000	0.002	0.156	0.000	0.003	0.198
Mean(y)	78.249	81.009	87.39	78.249	81.009	87.39
Panel C. Grade repetition rate						
Rural	-0.065	0.613	3.085*	-0.416	0.898	5.413*
	(0.635)	(1.424)	(1.612)	(1.476)	(2.736)	(2.796)
Urban	2.287***	1.264	2.58	10.152***	3.267	3.414
	(0.669)	(1.438)	(1.626)	(2.032)	(2.782)	(2.740)
P-value (rural=urban)	0.000	0.163	0.267	0.000	0.189	0.219
Mean(y)	9.373	6.54	4.499	9.373	6.54	4.499
Panel C. Dropout rate						
Rural	0.980*	0.739	2.467	2.071*	0.548	3.069
	(0.514)	(1.588)	(1.751)	(1.199)	(3.071)	(2.858)
Urban	2.761***	2.787*	4.638***	10.172***	8.156***	10.137***
	(0.592)	(1.582)	(1.791)	(1.966)	(2.981)	(2.933)
P-value (rural=urban)	0.000	0.000	0.001	0.000	0.000	0.002
Mean(y)	6.484	7.886	4.878	6.484	7.886	4.878
First-stage F	298.2	27.6	18.5	331.8	42.8	28.5
Observations	47,740	10,509	6,214	47,740	10,509	6,214

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate panel regression that controls for school and year fixed effects, and municipal time trends. The heterogeneous effects and the difference between groups are estimated using fully interacted models. IV regressions instrument the mining intensity measures (and their interaction with the urban/rural group) with the interaction between gold deposits in the neighborhood and international prices (and its interaction with the urban/rural group). The first-stage F corresponds to the Kleibergen-Paap Wald statistic. Standard errors are clustered at school level. The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2004-2012. A mine is considered in the neighborhood of a school if the distance is smaller or equal to 20 km. Active titles and mining deforestation, as defined in Section 3.3.2, are normalized with mean zero and standard deviation equal to one.

Table 3.4: Effect of Gold Mining on School Enrollment and Progress Throughout the Year: By Gender (IV only, 20 km Neighborhood)

	Active titles			Mining deforestation		
	Primary	Middle	High	Primary	Middle	High
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Enrollment (beginning of the year)						
Female	-0.573	0.058	3.689	-2.239	-4.798	7.592
	(1.096)	(4.683)	(3.455)	(2.582)	(9.162)	(5.728)
Male	0.3	4.666	2.152	1.602	13.673	2.012
	(1.114)	(4.595)	(3.429)	(2.661)	(8.872)	(5.648)
P-value (female=male)	0.001	0.000	0.022	0.001	0.000	0.021
Mean(y)	37.956	95.429	49.443	37.956	95.429	49.443
Panel B. Grade promotion rate						
Female	-2.473**	-1.954	-4.393*	-5.939**	-2.693	-5.403
	(0.979)	(2.441)	(2.671)	(2.331)	(4.485)	(4.326)
Male	-2.286**	-2.907	-6.057**	-5.116**	-6.518	-11.465***
	(0.975)	(2.441)	(2.676)	(2.319)	(4.471)	(4.299)
P-value (female=male)	0.389	0.004	0.000	0.389	0.004	0.000
Mean(y)	78.335	80.891	87.176	78.335	80.891	87.176
Panel C. Grade repetition rate						
Female	0.114	0.472	2.251	-0.211	-0.488	2.113
	(0.653)	(1.400)	(1.686)	(1.539)	(2.611)	(2.762)
Male	0.573	1.801	3.757**	1.809	4.831*	7.589***
	(0.659)	(1.409)	(1.719)	(1.567)	(2.631)	(2.863)
P-value (female=male)	0.003	0.000	0.000	0.003	0.000	0.000
Mean(y)	9.309	6.562	4.605	9.309	6.562	4.605
Panel C. Dropout rate						
Female	1.573***	2.204	3.641**	4.311***	4.392	5.646**
	(0.532)	(1.760)	(1.617)	(1.266)	(3.300)	(2.587)
Male	0.937*	1.974	3.859**	1.514	3.482	6.454**
	(0.534)	(1.737)	(1.595)	(1.271)	(3.218)	(2.510)
P-value (female=male)	0.000	0.325	0.493	0.000	0.332	0.484
Mean(y)	6.457	7.95	4.955	6.457	7.95	4.955
First stage F	298.1	27.5	19.0	329.6	42.4	28.8
Observations	95,480	21,018	12,428	95,480	21,018	12,428

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate panel regression that controls for school and year fixed effects, and municipal time trends. The heterogeneous effects and the difference between groups are estimated using fully interacted models. IV regressions instrument the mining intensity measures (and their interaction with the gender group) with the interaction between gold deposits in the neighborhood and international prices (and its interaction with the gender group). The first-stage F corresponds to the Kleibergen-Paap Wald statistic. Standard errors are clustered at school level. The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2004-2012. A mine is considered in the neighborhood of a school if the distance is smaller or equal to 20 km. Active titles and mining deforestation, as defined in Section 3.3.2, are normalized with mean zero and standard deviation equal to one.

Table 3.5: Effect of Gold Mining on Students Taking the Exit Exam (20 km Neighborhood)

	Active titles			Mining deforestation		
	Students per school (1)	Exam score (2)	Student works* (3)	Students per school (4)	Exam Score (5)	Student works* (6)
OLS	2.198	0.012	0.003	0.349	0.016**	0
	(1.766)	(0.008)	(0.004)	(0.854)	(0.006)	(0.003)
IV	8.097*	0.013	0.001	13.832*	0.011	0.001
	(4.433)	(0.010)	(0.003)	(7.439)	(0.007)	(0.003)
Mean(y)	50.83	0	0.129	50.83	0	0.129
First-stage F	37.5	220.1	554.1	61.9	120.2	308.4
Observations	5180	318226	195879	5180	318226	195879

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate regression that controls for school and year fixed effects, and municipal-specific time trends. Individual regressions also control for the students' age and gender. IV regressions instrument the mining intensity measures with the interaction between gold deposits in the neighborhood and international prices. Standard errors are clustered at school level. The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2004-2012, except for work situation that is restricted to 2008-2012. A mine is considered in the neighborhood of a school if the distance is smaller or equal to 20 km. Active titles and mining deforestation, as defined in Section 3.3.2, and test scores are normalized with mean zero and standard deviation equal to one.

Table 3.6: Effect of Gold Mining on Students Taking the Exit Exam: By Urban/Rural (IV only, 20 km Neighborhood)

	Active titles			Mining deforestation		
	Students per school (1)	Exam score (2)	Student works* (3)	Students per school (4)	Exam Score (5)	Student works* (6)
Rural	8.449*	0.011	0.001	17.438**	0.008	0.001
	(4.435)	(0.010)	(0.003)	(7.420)	(0.008)	(0.003)
Urban	5.987	0.014	0.002	5.848	0.012*	0.002
	(5.321)	(0.009)	(0.004)	(10.140)	(0.007)	(0.004)
P-value (rural=urban)	0.184	0.298	0.331	0.098	0.394	0.330
Mean(y)	50.830	0.000	0.129	50.830	0.000	0.129
First-stage F	25.8	66.9	163.0	18.1	107.7	263.1
Observations	5,180	318,226	195,879	5,180	318,226	195,879

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate regression that controls for school and year fixed effects, and municipal-specific time trends. Individual regressions also control for the students' age and gender. The heterogeneous effects and the difference between groups are estimated using fully interacted models. IV regressions instrument the mining intensity measures (and their interaction with the urban/rural group) with the interaction between gold deposits in the neighborhood and international prices (and its interaction with the urban/rural group). The first-stage F corresponds to the Kleibergen-Paap Wald statistic. Standard errors are clustered at school level. The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2004-2012, except for work situation that is restricted to 2008-2012. A mine is considered in the neighborhood of a school if the distance is smaller or equal to 20 km. Test scores, active titles and mining deforestation, as defined in Section 3.3.2, and test scores are normalized normalized with mean zero and standard deviation equal to one.

Table 3.7: Effect of Gold Mining on Students Taking the Exit Exam: By Gender (IV only, 20 km Neighborhood)

	Active titles			Mining deforestation		
	Students per school (1)	Exam score (2)	Student works* (3)	Students per school (4)	Exam Score (5)	Student works* (6)
Female	6.678**	0.019**	-0.010	12.824***	0.017**	-0.011
	(2.706)	(0.009)	(0.014)	(4.047)	(0.007)	(0.015)
Male	4.494*	0.009	0.016	4.187	0.006	0.017
	(2.606)	(0.010)	(0.014)	(3.649)	(0.009)	(0.015)
P-value (female=male)	0.000	0.175	0.185	0.000	0.177	0.185
Mean(y)	24.696	0.000	0.129	24.696	0.000	0.129
First stage F	35.2	82.9	202.4	22.4	146.3	357.7
Observations	12,458	308,871	195,879	12,458	308,871	195,879

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate regression that controls for school and year fixed effects, and municipal-specific time trends. Individual regressions also control for the students' age and gender. The heterogeneous effects and the difference between groups are estimated using fully interacted models. IV regressions instrument the mining intensity measures (and their interaction with the gender group) with the interaction between gold deposits in the neighborhood and international prices (and its interaction with the gender group). The first-stage F corresponds to the Kleibergen-Paap Wald statistic. Standard errors are clustered at school level. The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2004-2012, except for work situation that is restricted to 2008-2012. A mine is considered in the neighborhood of a school if the distance is smaller or equal to 20 km. Test scores, active titles and mining deforestation, as defined in Section 3.3.2, and test scores are normalized normalized with mean zero and standard deviation equal to one.

Table 3.8: Effect of Gold Mining on School Attendance and Child Labor (20 km Neighborhood)

	Active titles					Mining deforestation				
	All	6-8	9-11	12-14	15-17	All	6-8	9-11	12-14	15-17
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. School attendance										
OLS	-0.021	-0.046	-0.036**	-0.012	-0.052	0.027	-0.033	-0.003	0.052	0.044
	(0.033)	(0.059)	(0.016)	(0.056)	(0.074)	(0.044)	(0.101)	(0.025)	(0.061)	(0.091)
IV	-0.035	-0.044	-0.042**	-0.033	-0.083	-0.075	-0.105	-0.079**	-0.078	-0.165
	(0.035)	(0.067)	(0.018)	(0.062)	(0.077)	(0.077)	(0.157)	(0.037)	(0.146)	(0.159)
Mean(y)	0.838	0.647	0.977	0.938	0.779	0.838	0.647	0.977	0.938	0.779
Panel B. Works										
OLS	0.008	0.002	0.044**	0.017	-0.034	0.024	0.009	0.049	0.041	-0.002
	(0.022)	(0.001)	(0.021)	(0.033)	(0.074)	(0.033)	(0.007)	(0.040)	(0.055)	(0.104)
IV	0.025	0	0.050**	-0.01	0.037	0.054	0.001	0.094**	-0.023	0.074
	(0.023)	(0.001)	(0.023)	(0.033)	(0.077)	(0.049)	(0.003)	(0.048)	(0.077)	(0.152)
Mean(y)	0.08	0.008	0.029	0.082	0.187	0.08	0.008	0.029	0.082	0.187
First-stage F	76.9	59.2	101.5	54.3	51.7	274.0	117.3	268.6	180.2	284.2
Observations	3,628	846	868	963	951	3,628	846	868	963	951

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate regression that controls for municipal and year fixed effects, and students characteristics (age, gender, household sized and parents' education). IV regressions instrument the mining intensity measures with the interaction between gold deposits in the neighborhood and international prices. Standard errors are clustered at DHS cluster level. The sample includes all urban households in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2005-2010. A mine is considered in the neighborhood of a cluster if the distance is smaller or equal to 20 km. Active titles and mining deforestation, as defined in Section 3.3.2, are normalized with mean zero and standard deviation equal to one.

Table 3.9: Effect of Gold Mining on School Attendance
and Child Labor, by Parents' Education (IV only, 20 km Neighborhood)

	Active titles					Mining deforestation				
	All	6-8	9-11	12-14	15-17	All	6-8	9-11	12-14	15-17
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. School attendance										
Primary education	0.064	0.024	0.064	0.074	-0.075	0.152	0.032	0.146	0.216	-0.138
	(0.053)	(0.152)	(0.058)	(0.082)	(0.092)	(0.125)	(0.270)	(0.140)	(0.236)	(0.222)
Secondary or higher	-0.077*	-0.054	-0.038	-0.104	-0.113	-0.175*	-0.146	-0.076	-0.329	-0.226
	(0.041)	(0.069)	(0.029)	(0.071)	(0.097)	(0.100)	(0.183)	(0.061)	(0.227)	(0.197)
P-value (primary=higher)	0.015	0.593	0.130	0.014	0.589	0.030	0.528	0.171	0.054	0.592
Mean(y)	0.838	0.647	0.977	0.938	0.779	0.838	0.647	0.977	0.938	0.779
Panel B. Works										
Primary education	0.018	-0.022	-0.009	-0.057	0.091	0.032	-0.036	-0.041	-0.152	0.206
	(0.043)	(0.014)	(0.044)	(0.062)	(0.097)	(0.092)	(0.027)	(0.101)	(0.169)	(0.230)
Secondary or higher	0.029	0.004	0.061**	0.018	0.036	0.066	0.012	0.122*	0.07	0.069
	(0.024)	(0.004)	(0.029)	(0.037)	(0.096)	(0.053)	(0.012)	(0.065)	(0.115)	(0.193)
P-value (primary=higher)	0.895	0.098	0.134	0.343	0.667	0.839	0.144	0.156	0.384	0.667
Mean(y)	0.08	0.008	0.029	0.082	0.187	0.08	0.008	0.029	0.082	0.187
First-stage F	38.1	12.0	16.6	43.0	56.3	7.6	3.4	4.7	5.0	8.8
Observations	3,627	846	868	963	950	3,627	846	868	963	950

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate regression that controls for municipal and year fixed effects, and students characteristics (age, gender, household sized and parents' education). The heterogeneous effects and the difference between groups are estimated using fully interacted models. IV regressions instrument the mining intensity measures (and their interaction with the parents' education group) with the interaction between gold deposits in the neighborhood and international prices (and its interaction with the parents' education group). The first-stage F corresponds to the Kleibergen-Paap Wald statistic. Standard errors are clustered at DHS cluster level. The sample includes all urban households in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2005-2010. A mine is considered in the neighborhood of a cluster if the distance is smaller or equal to 20 km. Active titles and mining deforestation, as defined in Section 3.3.2, are normalized with mean zero and standard deviation equal to one.

Table 3.10: Effect of Gold Mining on School Attendance and Child Labor, by Gender (IV only, 20 km Neighborhood)

	Active titles					Mining deforestation				
	All	6-8	9-11	12-14	15-17	All	6-8	9-11	12-14	15-17
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. School attendance										
Female	-0.07	-0.063	-0.027	-0.099	-0.119	-0.191	-0.158	-0.04	-0.399	-0.252
	(0.045)	(0.073)	(0.046)	(0.084)	(0.081)	(0.121)	(0.180)	(0.139)	(0.341)	(0.177)
Male	-0.009	0.056	-0.050**	-0.003	-0.038	0.003	0.174	-0.094**	0.021	-0.02
	(0.037)	(0.106)	(0.021)	(0.061)	(0.080)	(0.090)	(0.336)	(0.040)	(0.157)	(0.214)
P-value (female=male)	0.160	0.290	0.551	0.085	0.302	0.184	0.373	0.679	0.183	0.354
Mean(y)	0.838	0.647	0.977	0.938	0.779	0.838	0.647	0.977	0.938	0.779
Panel B. Works										
Female	0.019	0.002	-0.031	0.014	0.034	0.039	0.005	-0.165	0.075	0.072
	(0.036)	(0.003)	(0.070)	(0.064)	(0.086)	(0.099)	(0.008)	(0.290)	(0.265)	(0.178)
Male	0.021	-0.006	0.085**	-0.012	0.013	0.047	-0.019	0.162*	-0.034	0.011
	(0.030)	(0.006)	(0.038)	(0.037)	(0.095)	(0.076)	(0.020)	(0.084)	(0.097)	(0.261)
P-value (female=male)	0.957	0.316	0.135	0.727	0.831	0.957	0.348	0.315	0.726	0.847
Mean(y)	0.08	0.008	0.029	0.082	0.187	0.08	0.008	0.029	0.082	0.187
First-stage F	22.2	16.5	16.5	19.9	27.7	4.1	2.8	3.0	2.0	4.7
Observations	3,627	846	868	963	950	3,627	846	868	963	950

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate regression that controls for municipal and year fixed effects, and students characteristics (age, gender, household sized and parents' education). The heterogeneous effects and the difference between groups are estimated using fully interacted models. IV regressions instrument the mining intensity measures (and their interaction with the gender group) with the interaction between gold deposits in the neighborhood and international prices (and its interaction with the gender group). The first-stage F corresponds to the Kleibergen-Paap Wald statistic. Standard errors are clustered at DHS cluster level. The sample includes all urban households in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2005-2010. A mine is considered in the neighborhood of a cluster if the distance is smaller or equal to 20 km. Active titles and mining deforestation, as defined in Section 3.3.2, are normalized with mean zero and standard deviation equal to one.

Table 3.11: Effect of Active Titles on School Enrollment and Progress Throughout the Year, by Distance (IV only)

	10 *km* (1)	20 *km* (2)	30 *km* (3)	40 *km* (4)	50 *km* (5)
Panel A. Enrollment (beginning of the year)					
Primary	2.05	-0.274	-0.954	-3.69	-2.423
	(3.010)	(2.193)	(2.770)	(2.257)	(3.080)
Middle	4.708	4.725	-0.961	10.324	-7.311
	(4.419)	(9.218)	(10.400)	(10.607)	(11.967)
High	1.909	5.841	-0.418	11.341	5.183
	(4.550)	(6.852)	(9.190)	(8.440)	(11.436)
Panel B. Grade promotion rate					
Primary	-1.397***	-2.322**	-2.839*	-0.25	-0.537
	(0.540)	(0.941)	(1.537)	(1.258)	(1.360)
Middle	0.089	-1.472	-1.343	0.23	3.152
	(1.169)	(2.303)	(2.556)	(2.826)	(3.330)
High	-0.703	-4.795*	-10.484***	-6.973*	-3.025
	(1.766)	(2.651)	(3.719)	(3.669)	(4.775)
Panel C. Grade repetition rate					
Primary	0.788**	0.283	0.336	-0.36	0.116
	(0.389)	(0.634)	(1.021)	(0.902)	(1.036)
Middle	-0.795	0.844	0.747	2.245	-0.235
	(0.758)	(1.416)	(1.887)	(1.993)	(2.330)
High	-0.082	2.913*	6.166***	5.842**	0.884
	(0.912)	(1.611)	(2.371)	(2.281)	(3.493)
Panel D. Dropout rate					
Primary	0.529*	1.244**	1.305	0.362	0.124
	(0.277)	(0.511)	(0.836)	(0.668)	(0.753)
Middle	0.865	1.465	0.337	-2.129	-2.462
	(0.809)	(1.563)	(1.702)	(1.774)	(1.975)
High	1.845*	3.207**	3.003	1.044	1.136
	(1.093)	(1.634)	(2.054)	(2.097)	(2.512)

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate instrumental variable panel regression that controls for school and year fixed effects, and municipal time trends. IV regressions instrument the mining intensity measures with the interaction between gold deposits in the neighborhood and international prices. Standard errors are clustered at school level. The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2004-2012. Each column corresponds to an alternative definition of neighborhood based on distances between mines and schools ranging from 10 to 50 km. Active titles, as defined in Section 3.3.2, are normalized with mean zero and standard deviation equal to one.

Table 3.12: Effect of Mining Deforestation on School Enrollment and Progress Throughout the Year, by Distance (IV only)

	10 *km* (1)	20 *km* (2)	30 *km* (3)	40 *km* (4)	50 *km* (5)
Panel A. Enrollment (beginning of the year)					
Primary	8.635	-0.637	-1.604	-6.403	-4.057
	(12.726)	(5.103)	(4.657)	(3.922)	(5.158)
Middle	25.467	8.876	-1.485	12.681	-8.179
	(25.447)	(17.497)	(16.072)	(12.984)	(13.386)
High	9.034	9.604	-0.514	12.031	4.697
	(21.766)	(11.116)	(11.308)	(8.946)	(10.345)
Panel B. Grade promotion rate					
Primary	-5.884**	-5.402**	-4.773*	-0.434	-0.899
	(2.311)	(2.205)	(2.606)	(2.184)	(2.278)
Middle	0.48	-2.766	-2.074	0.283	3.526
	(6.330)	(4.226)	(3.931)	(3.471)	(3.704)
High	-3.328	-7.884*	-12.896***	-7.398*	-2.741
	(8.609)	(4.292)	(4.645)	(3.896)	(4.345)
Panel C. Grade repetition rate					
Primary	3.318**	0.659	0.564	-0.624	0.194
	(1.648)	(1.476)	(1.720)	(1.563)	(1.735)
Middle	-4.297	1.585	1.154	2.758	-0.263
	(4.386)	(2.618)	(2.905)	(2.444)	(2.605)
High	-0.39	4.790*	7.585**	6.198**	0.801
	(4.327)	(2.665)	(3.077)	(2.443)	(3.167)
Panel D. Dropout rate					
Primary	2.230*	2.895**	2.194	0.628	0.207
	(1.182)	(1.195)	(1.402)	(1.160)	(1.261)
Middle	4.68	2.753	0.52	-2.615	-2.754
	(4.698)	(2.893)	(2.630)	(2.159)	(2.186)
High	8.729	5.273**	3.694	1.108	1.029
	(6.995)	(2.593)	(2.530)	(2.223)	(2.279)

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate instrumental variable panel regression that controls for school and year fixed effects, and municipal time trends. IV regressions instrument the mining intensity measures with the interaction between gold deposits in the neighborhood and international prices. Standard errors are clustered at school level. The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2004-2012. Each column corresponds to an alternative definition of neighborhood based on distances between mines and schools ranging from 10 to 50 km. Mining deforestation, as defined in Section 3.3.2, is normalized with mean zero and standard deviation equal to one.

Table 3.13: Effect of Gold Mining on Students Taking the Exit Exam, By Distance (IV only)

	10 *km* (1)	20 *km* (2)	30 *km* (3)	40 *km* (4)	50 *km* (5)
Panel A. Active titles					
Students per school	12.064	13.832*	-0.104	-4.11	-0.141
	(11.124)	(7.744)	(7.072)	(6.041)	(6.906)
Exam score	0.009	0.011	0.016*	0.019**	0.014
	(0.007)	(0.007)	(0.009)	(0.009)	(0.010)
Student works	-0.001	0.001	0.005	0.006	0.009
	(0.003)	(0.003)	(0.005)	(0.007)	(0.008)
Panel A. Mining deforestation					
Students per school	3.466	8.097*	-0.109	-4.135	-0.126
	(3.105)	(4.558)	(7.380)	(6.035)	(6.157)
Exam score	0.01	0.013	0.027*	0.038**	0.03
	(0.008)	(0.010)	(0.014)	(0.018)	(0.022)
Student works	-0.001	0.001	0.005	0.006	0.009
	(0.003)	(0.003)	(0.005)	(0.007)	(0.008)

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate regression that controls for school and year fixed effects, and municipal-specific time trends. Individual regressions also control for the students' age and gender. IV regressions instrument the mining intensity measures with the interaction between gold deposits in the neighborhood and international prices. Standard errors are clustered at school level. The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2004-2012, except for work situation that is restricted to 2008-2012. Each column corresponds to an alternative definition of neighborhood based on distances between mines and schools ranging from 10 to 50 km. Test scores, active titles and mining deforestation, as defined in Section 3.3.2, and test scores are normalized with mean zero and standard deviation equal to one.

Table 3.14: Effect of Active Titles on School Attendance and Child Labor,
by Distance (IV only)

	10 *km* (1)	20 *km* (2)	30 *km* (3)	40 *km* (4)	50 *km* (5)
Panel A. School attendance					
All	-0.012	-0.035	-0.051**	-0.051*	-0.045
	(0.021)	(0.035)	(0.024)	(0.029)	(0.029)
6-8	-0.033	-0.044	-0.007	-0.005	-0.023
	(0.035)	(0.067)	(0.044)	(0.050)	(0.049)
9-11	-0.021**	-0.042**	-0.035*	-0.044*	-0.053**
	(0.009)	(0.018)	(0.019)	(0.025)	(0.023)
12-14	-0.009	-0.033	-0.057	-0.072	-0.068
	(0.031)	(0.062)	(0.054)	(0.064)	(0.062)
15-17	-0.011	-0.083	-0.098	-0.091	-0.057
	(0.045)	(0.077)	(0.065)	(0.082)	(0.084)
Panel B. Works					
All	0.002	0.025	0.023	0.022	0.018
	(0.014)	(0.023)	(0.017)	(0.021)	(0.022)
6-8	0.000	0.000	-0.001	0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
9-11	0.023*	0.050**	0.047**	0.061***	0.062***
	(0.012)	(0.023)	(0.019)	(0.020)	(0.020)
12-14	0.009	-0.01	-0.011	-0.006	0.000
	(0.018)	(0.033)	(0.027)	(0.033)	(0.032)
15-17	-0.027	0.037	0.046	0.022	-0.007
	(0.041)	(0.077)	(0.064)	(0.081)	(0.083)

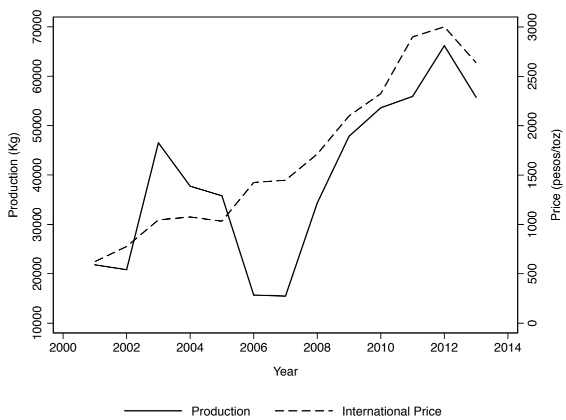
Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate regression that controls for municipal and year fixed effects, and students characteristics (age, gender, household sized and parents' education). IV regressions instrument the mining intensity measures with the interaction between gold deposits in the neighborhood and international prices. Standard errors are clustered at DHS cluster level. The sample includes all urban households in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2005-2010. Each column corresponds to an alternative definition of neighborhood based on distances between mines and clusters ranging from 10 to 50 km. Active titles, as defined in Section 3.3.2, is normalized with mean zero and standard deviation equal to one.

Table 3.15: Effect of Mining Deforestation on School Attendance and Child Labor,
by Distance (IV only)

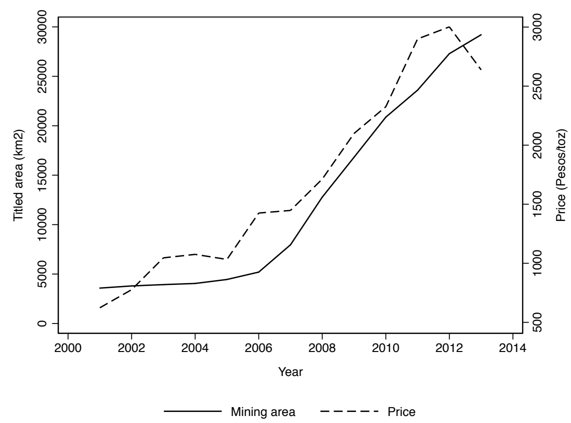
	10 *km* (1)	20 *km* (2)	30 *km* (3)	40 *km* (4)	50 *km* (5)
Panel A. School attendance					
All	-0.057	-0.075	-0.174**	-0.132*	-0.105
	(0.102)	(0.077)	(0.085)	(0.077)	(0.070)
6-8	-0.22	-0.105	-0.022	-0.012	-0.061
	(0.293)	(0.157)	(0.141)	(0.131)	(0.129)
9-11	-0.077**	-0.079**	-0.100*	-0.106*	-0.110**
	(0.035)	(0.037)	(0.057)	(0.063)	(0.053)
12-14	-0.038	-0.078	-0.213	-0.186	-0.163
	(0.136)	(0.146)	(0.199)	(0.165)	(0.149)
15-17	-0.052	-0.165	-0.317	-0.227	-0.118
	(0.210)	(0.159)	(0.217)	(0.209)	(0.176)
Panel B. Works					
All	0.01	0.054	0.08	0.058	0.041
	(0.065)	(0.049)	(0.055)	(0.055)	(0.052)
6-8	0.000	0.001	-0.002	-0.001	0.002
	(0.006)	(0.003)	(0.004)	(0.002)	(0.002)
9-11	0.082*	0.094**	0.134***	0.146***	0.130***
	(0.043)	(0.048)	(0.051)	(0.050)	(0.048)
12-14	0.04	-0.023	-0.04	-0.017	0.000
	(0.082)	(0.077)	(0.104)	(0.086)	(0.077)
15-17	-0.126	0.074	0.148	0.054	-0.015
	(0.203)	(0.152)	(0.206)	(0.201)	(0.171)

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate regression that controls for municipal and year fixed effects, and students characteristics (age, gender, household sized and parents' education). IV regressions instrument the mining intensity measures with the interaction between gold deposits in the neighborhood and international prices. Standard errors are clustered at DHS cluster level. The sample includes all urban households in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2005-2010. Each column corresponds to an alternative definition of neighborhood based on distances between mines and clusters ranging from 10 to 50 km. Mining deforestation, as defined in Section 3.3.2, is normalized with mean zero and standard deviation equal to one.

Figure 3.1: The Colombian Gold Rush



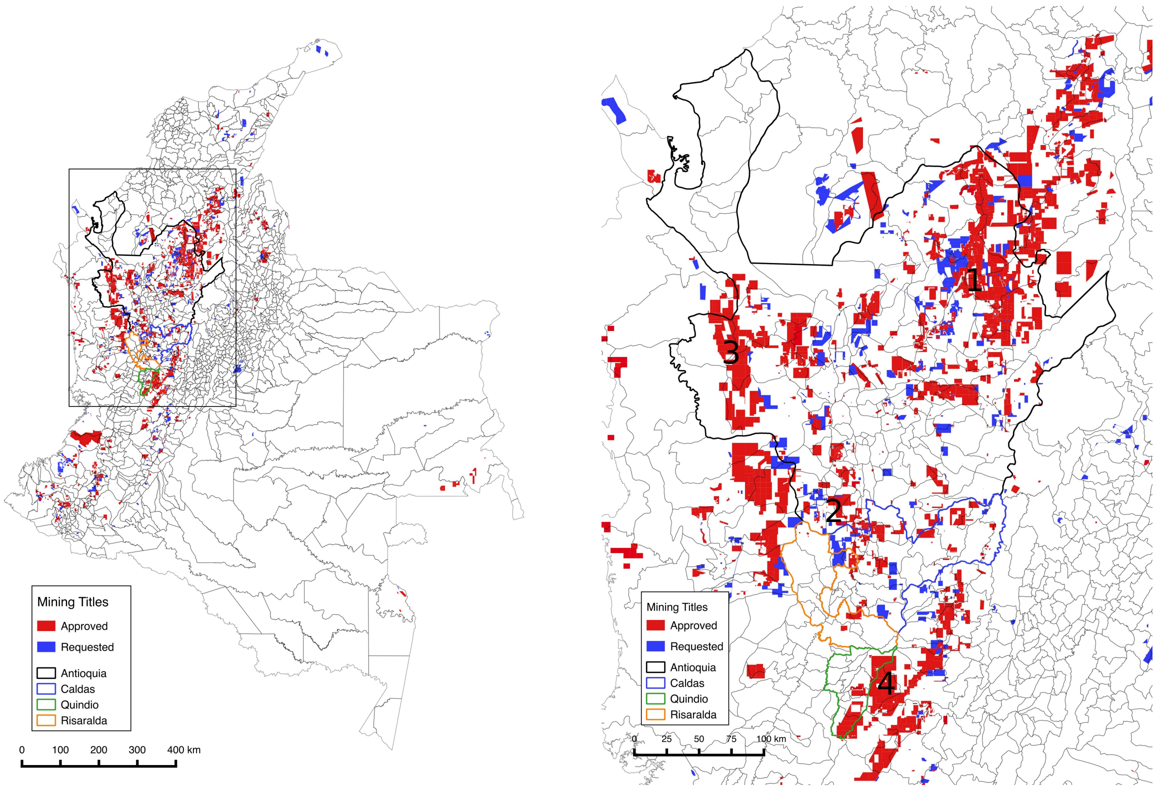
(a) Reported Prodution



(b) Approved Mining Titles

Source: Own calculations based on (a) SIMCO; (b) SIMCO, SIGOT and Tierraminada.
Notes: Gold production is expressed in kg and titled area in km^2. International prices are expressed in Colombian pesos per Troy Once.

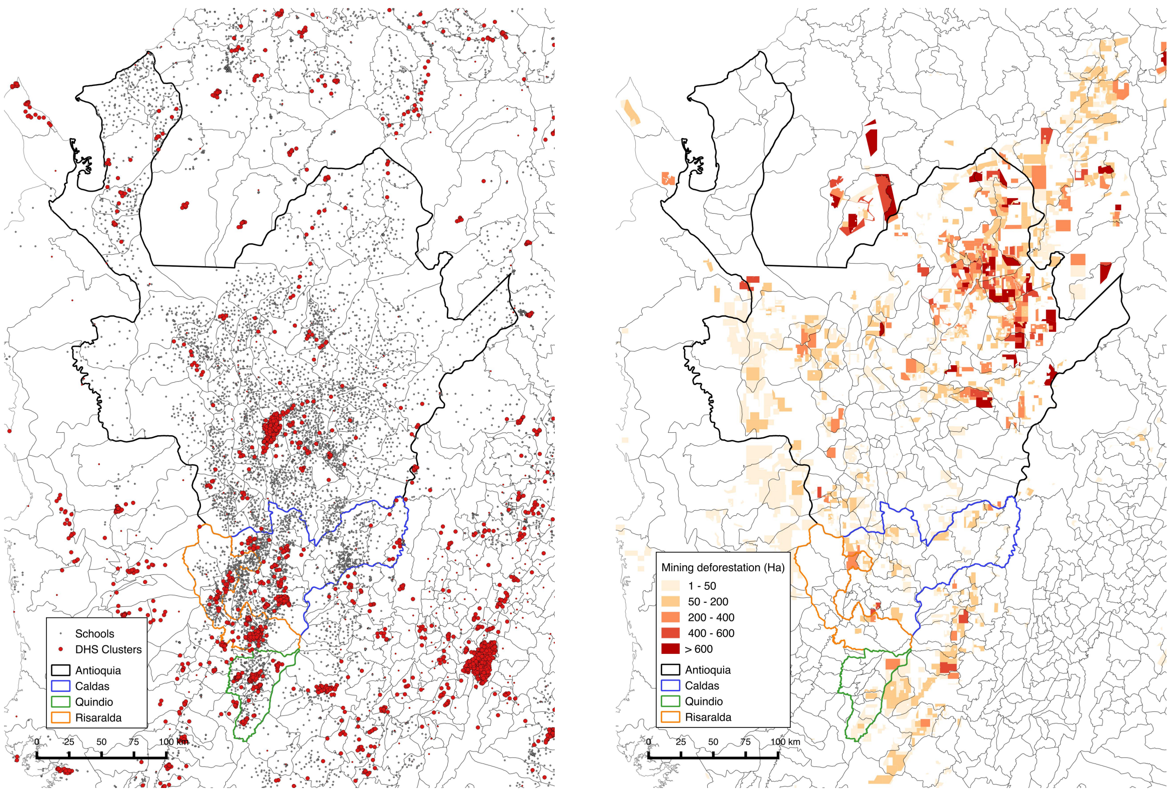
Figure 3.2: Gold Mining Titles



(a) Colombia **(b)** Antioquia and the Coffee Region

Source: Own calculations based on SIMCO, SIGOT and Tierraminada;
Notes: Approved titles include all exploration and exploitation titles that were active at some point between 2000 and 2014. Requested titles include all pending requests in November 2014.

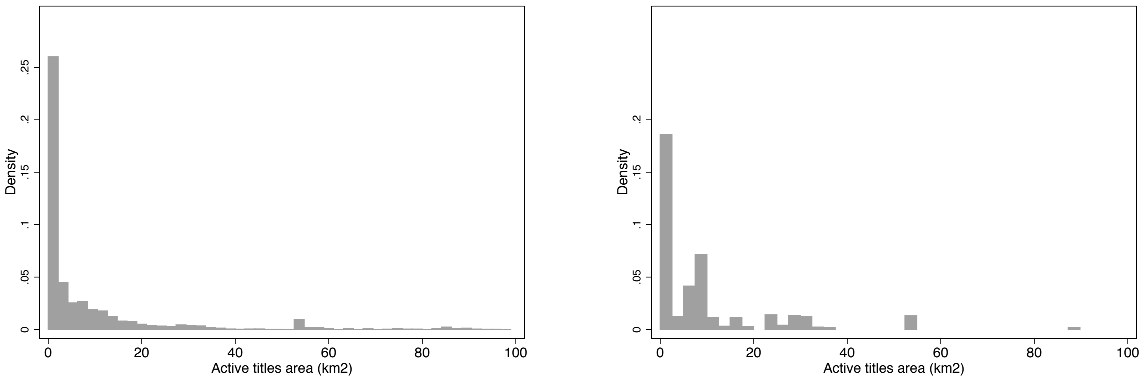
Figure 3.3: Schools, DHS clusters and Mining Deforestation



(a) Schools and DHS Clusters **(b)** Mining Deforestation (2001-2012)

Source: Own calculations based on (a) Ministry of Education and DHS; (b) SIMCO, SIGOT, Tierraminada and Hansen et al. (2013).
Notes: (a) 2005 DHS clusters, and missing urban schools are located using the coordinates of the corresponding municipal town. (b) The color scale represents the total mining deforestation between 2001 and 2012, expressed in Ha.

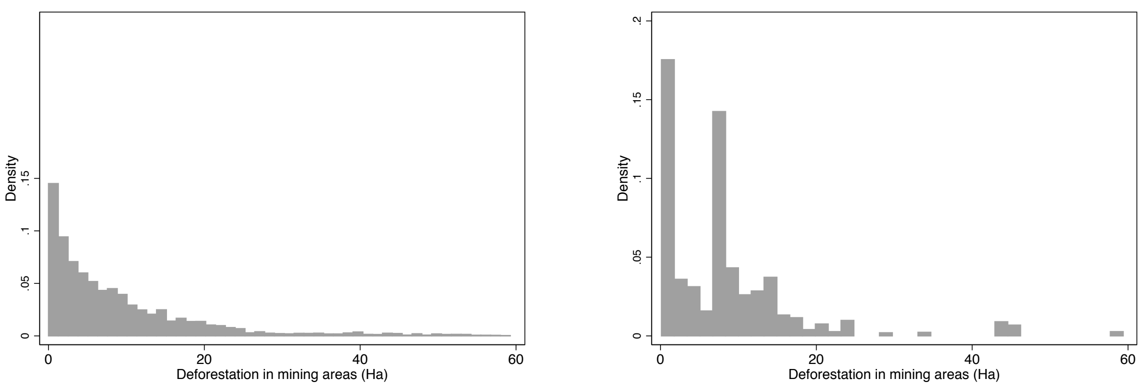
Figure 3.4: Active Mines in the Neighborhood (20 km)



(a) Schools **(b)** DHS Urban Clusters

Source: Own calculations based on Ministry of Education, DHS, SIMCO, SIGOT and Tierraminada.
Notes: The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. A mine is considered in the neighborhood of a school if the distance is smaller or equal to 20 km. Active titles are defined in Section 3.3.2.

Figure 3.5: Mining Deforestation in the Neighborhood (20 km)



(a) Schools **(b)** DHS Urban Clusters

Source: Own calculations based on Ministry of Education, DHS, SIMCO, SIGOT, Tierraminada and Hansen et al. (2013).
Notes: The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. A mine is considered in the neighborhood of a school if the distance is smaller or equal to 20 km. Mining deforestation is defined in Section 3.3.2.

Bibliography

Acemoglu, D., García-Jimeno, C., and Robinson, J. A. (2012). Finding eldorado: Slavery and long-run development in colombia. *Journal of Comparative Economics*, 40(4):534–564. 81

Aguilera, J. A. R. et al. (2014). Sobre el efecto de las regalías en el bienestar: Una revisión del periodo 2001-2011. Technical report, Departamento Nacional de Planeación. 100

Alvarez-Berríos, N. L. and Aide, T. M. (2015). Global demand for gold is another threat for tropical forests. *Environmental Research Letters*, 10(1):014006. 79

Angelucci, M. and De Giorgi, G. (2009). Indirect effects of an aid program: how do cash transfers affect ineligibles' consumption? *The American Economic Review*, pages 486–508. 49

Angrist, J. D. and Lang, K. (2004). Does school integration generate peer effects? evidence from boston's metco program. *The American Economic Review*, 94(5):1613–1634. 49

Appadurai, A. (2004). The capacity to aspire: Culture and the terms of recognition. In Rao, V. and Walton, M., editors, *Culture and Public Action*. Stanford University Press. 26

Aragón, F. M. and Rud, J. P. (2013). Natural resources and local communities: evidence from a peruvian gold mine. *American Economic Journal: Economic Policy*, 5(2):1–25. 76

Attanasio, O. and Kaufmann, K. (2009). Educational Choices, Subjective Expectations, and Credit Constraints. NBER Working Papers 15087, National Bureau of Economic Research, Inc. 2

Avitabile, C. and De Hoyos Navarro, R. E. (2015). The Heterogeneous Effect of Information on Student Performance: Evidence from a Randomized Control Trial in Mexico. *World Bank Policy Research Working Paper*, (7422). 5

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In *ICWSM*, pages 361–362. 72

Baur, D. G. and McDermott, T. K. (2010). Is gold a safe haven? international evidence. *Journal of Banking & Finance*, 34(8):1886–1898. 78

Belzil, C. and Hansen, J. (2004). Earnings dispersion, risk aversion and education. In Polachek, S. W., editor, *Accounting for Worker Well-Being (Research in Labor Economics, Volume 23)*, pages 335–358. Emerald Group Publishing Limited. 26

Belzil, C. and Leonardi, M. (2007). Can risk aversion explain schooling attainments? Evidence from Italy. *Labour Economics*, 14(6):957–970. 26

Bénabou, R. and Tirole, J. (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics*, pages 871–915. 26

Bloomberg (2015). Gold ceo hunted in $970 million colombia laundering case. `http://www.bloomberg.com/news/articles/2015-01-16/colombian-gold-traders-arrested-in-970m-laundering-case`. Accessed: 2016-01-20. 80

Bobonis, G. J. and Finan, F. (2009). Neighborhood peer effects in secondary school enrollment decisions. *The Review of Economics and Statistics*, 91(4):695–716. 49

Booij, A. S., Leuven, E., and Oosterbeek, H. (2012). The role of information in the take-up of student loans. *Economics of Education Review*, 31(1):33–44. 2, 5, 25

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of econometrics*, 150(1):41–55. 50, 58

Burgess, S., Umaña-Aponte, M., et al. (2011). *Raising your sights: the impact of friendship networks on educational aspirations*. Centre for Market and Public Organisation. 46, 48, 52

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*. 26

Calvó-Armengol, A., Patacchini, E., and Zenou, Y. (2009). Peer effects and social networks in education. *The Review of Economic Studies*, 76(4):1239–1267. 50, 51

Campaign, C. S. (2013). La colosa: a death foretold alternative report about the anglogold ashanti gold mining project in cajamarca, tolima, colombia. 81

Cárdenas, J. C. (2014). Two tales of mining and human choice. *ReVista (Cambridge)*, 13(2):2. 81

Chuhan-Pole, P., Dabalen, A., Kotsadam, A., Sanoh, A., and Tolonen, A. K. (2015). The local socioeconomic effects of gold mining: evidence from ghana. *World Bank Policy Research Working Paper*, (7250). 76

Conti, G., Galeotti, A., Mueller, G., and Pudney, S. (2013). Popularity. *Journal of Human Resources*, 48(4):1072–1094. 51

Cordy, P., Veiga, M. M., Salih, I., Al-Saadi, S., Console, S., Garcia, O., Mesa, L. A., Velásquez-López, P. C., and Roeser, M. (2011). Mercury contamination from artisanal gold mining in antioquia, colombia: The world's highest per capita mercury pollution. *Science of the Total Environment*, 410:154–160. 73, 80

Dale, S. and Krueger, A. B. (2011). Estimating the return to college selectivity over the career using administrative earnings data. Technical report, National Bureau of Economic Research. 45

Dale, S. B. and Krueger, A. B. (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *Quarterly Journal of Economics*, 117(4). 45

Dalton, P. S., Ghosal, S., and Mani, A. (2016). Poverty and aspirations failure. *The Economic Journal*, 126(590):165–188. 26

De Giorgi, G., Pellizzari, M., and Redaelli, S. (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, pages 241–275. 46, 48, 50, 52

Defensoría del Pueblo (2010). La minería de hecho en colombia. *Bogota: Defensoria del Pueblo*. 74, 79

Del Bello, C. L., Patacchini, E., and Zenou, Y. (2015). Neighborhood effects in education. 64

Deming, D. J., Hastings, J. S., Kane, T. J., and Staiger, D. O. (2014). School choice, school quality and postsecondary attainment. Technical Report 3. 50

Ding, W. and Lehrer, S. F. (2007). Do peers affect student achievement in china's secondary schools? *The Review of Economics and Statistics*, 89(2):300–312. 50

Dinkelman, T. and Martínez, C. (2014). Investing in Schooling In Chile: The Role of Information about Financial Aid for Higher Education. *The Review of Economics and Statistics*, 96(2):244–257. 2, 5, 25

DPLF, D. (2015). The right of indigenous peoples to prior consultation: The situation in bolivia, colombia, ecuador, and peru. 88

Dube, O. and Vargas, J. F. (2013). Commodity price shocks and civil conflict: Evidence from colombia. *The Review of Economic Studies*, 80(4):1384–1421. 74, 80

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review*, 101(5):1739–74. 50

Echavarria, C. (2014). What is legal? formalising artisanal and small-scale mining in colombia. *IIED, London and ARM, Colombia*. 79

Echeverry, J., Masmela, G., and García, A. (2011). Por qué es necesaria la creación de un sistema general de regalías. *Notas Fiscales, Ministerio de Hacienda y Crédito Público*, (2):459–483. 100

El Colombiano (2012). 50% del oro no paga impuesto de renta. `http://www.elcolombiano.com/historico/50_del_oro_no_paga_impuesto_de_renta-LFEC_204087`. Accessed: 2016-01-20. 80

El Espectador (2013a). La ruta de las regalías ficticias. `http://www.elespectador.com/noticias/investigacion/ruta-de-regalias-ficticias-articulo-440663`. Accessed: 2016-01-20. 80

El Espectador (2013b). Tatequieto a la minería. `http://www.elespectador.com/noticias/judicial/tatequieto-mineria-articulo-472294`. Accessed: 2016-01-20. 74

El Tiempo (2013). Campamentos de explotación de niñas en zonas mineras. `http://www.eltiempo.com/archivo/documento/CMS-12824463`. Accessed: 2016-01-20. 74

Ellwood, D. and Kane, T. (2000). Who is getting a college education? Family background and the growing gaps in enrollment. In Danziger, S. and Waldfogel, J., editors, *Securing the Future*, pages 283–324. New York: Russell Sage Foundation. 1

Fletcher, J. M. (2015). Social interactions and college enrollment: A combined school fixed effects/instrumental variables approach. *Social science research*, 52:494–507. 46, 48, 51

Fryer, R. and Torelli, P. (2010). An empirical analysis of acting white. *Journal of Public Economics*, 94(5):380–396. 50

Fryer, R. G. (2013). Information and Student Achievement: Evidence from a Cellular Phone Experiment. NBER Working Papers 19113, National Bureau of Economic Research, Inc. 25

Gamboa, L. F. and Rodríguez, P. A. (2014). Do Colombian students underestimate higher education returns? *Working Paper 164. Universidad del Rosario*. 20

Garay, L. J. et al. (2013). Minería en colombia. fundamentos para superar el modelo extractivista. 80

Genicot, G. and Ray, D. (2014). Aspirations and inequality. Working Paper 19976, National Bureau of Economic Research. 26

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3):1049–1074. 26

Goldin, C., Katz, L. F., and Kuziemko, I. (2006). The Homecoming of American College Women: The Reversal of the Gender Gap in College. *Journal of Economic Perspectives*, 20:133–156. 26

Goldsmith-Pinkham, P. and Imbens, G. W. (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, 31(3):253–264. 51, 64

Goñi, E. A., Sabogal, A., and Asmat, R. (2014). Minería informal aurífera en colombia. 74, 79, 88, 94

Gonzalez, L. et al. (2013). Impacto de la minería de hecho en colombia. *Estudios de caso: Quibdó, Istmina, Timbiquí, López de Micay, Guapi, El Charco y Santa Bárbara. Instituto de Estudios para el De sarrollo y la Paz–INDEPAZ. Bogotá, Colombia*. 74, 79, 88

González-Velosa, C., Rucci, G., Sarzosa, M., and Urzúa, S. (2015). Returns to Higher Education in Chile and Colombia. Technical report, Inter-American Development Bank. 8

Griffith, A. L. and Rask, K. N. (2014). Peer effects in higher education: A look at heterogeneous impacts. *Economics of Education Review*, 39:65–77. 50

Güiza, L. and Aristizabal, J. D. (2013). Mercury and gold mining in colombia: a failed state. *Universitas Scientiarum*, 18(1):33–49. 80

Gylfason, T. (2001). Natural resources, education, and economic development. *European economic review*, 45(4):847–859. 77

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., Thau, D., Stehman, S., Goetz, S., Loveland, T., et al. (2013). High-resolution global maps of 21st-century forest cover change. *science*, 342(6160):850–853. Data available online: `http://earthenginepartners.appspot.com/science-2013-global-forest`. Accessed: 2015-04-23. 87, 117, 118, 145

Hanushek, E. A., Kain, J. F., Markman, J. M., and Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of applied econometrics*, 18(5):527–544. 50

Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2002). New evidence about brown v. board of education: The complex effects of school racial composition on achievement. Technical report, National Bureau of Economic Research. 49

Hastings, J., Neilson, C. A., and Zimmerman, S. D. (2015). The Effects of Earnings Disclosure on College Enrollment Decisions. NBER Working Papers 21300, National Bureau of Economic Research, Inc. 2, 3, 5, 6, 20, 25, 31

Hastings, J. S., Neilson, C. A., and Zimmerman, S. D. (2013). Are some degrees worth more than others? Evidence from college admission cutoffs in Chile. Technical report, National Bureau of Economic Research. 2, 45

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pages 153–161. 51

Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the national Academy of Sciences*, 104(33):13250–13255. 26

Heckman, J. J., Stixrud, J., and Urzúa, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482. 26

Heifetz, A. and Minelli, E. (2014). Aspiration traps. *The B.E. Journal of Theoretical Economics*, 15(2):125–142. 26

Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The Review of Economics and Statistics*, 91(4):717–724. 45

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. Technical report, National Bureau of Economic Research. 49, 50

Hoxby, C. and Turner, S. (2013). Expanding college opportunities for high-achieving, low income students. *Stanford Institute for Economic Policy Research Discussion Paper*, (12-014). 25, 31, 45

Hoxby, C. and Turner, S. (2015). What high-achieving low-income students know about college. Technical report, National Bureau of Economic Research. 2

Hsieh, C.-S. and Lee, L. F. (2014). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics*. 51, 52, 57, 59, 60, 61, 63

ICBF, I. (2001). Situación del niño minero en los municipios de istmina, condoto, tadó y río iró. 74

IDEAM, I. (2015). Informe del estado del medio ambiente y de los recursos naturales renovables. 73, 80, 87

Idrobo, N., Mejía, D., and Tribín, A. (2013). Minería ilegal y violencia en colombia. *CESED, Facultad de Economía, U. de los Andes. Bogotá-Colombia*. 74, 80

ILO, I., IPEC, I., , and MINERCOL (2001a). The boys and girls who work in colombia's small scale mining: Socio cultural economic and legislative diagnosis. 74

ILO, I., IPEC, I., , and MINERCOL (2001b). El trabajo infantil en la minería artesanal del oro. diagnóstico sociocultural y económico del municipio de condoto en chocó. 74

Imberman, S. A., Kugler, A. D., and Sacerdote, B. I. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *The American Economic Review*, 102(5):2048–2082. 50

Jensen, R. (2010). The (Perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, 125(2):515–548. 2, 5, 21

Kane, T. (1994). College entry by blacks since 1970: The role of college costs, family background, and the returns to education. *Journal of Political Economy*, 102(5):878–911. 1

Kaufmann, K. M. (2014). Understanding the income gradient in college attendance in mexico: The role of heterogeneity in expected returns. *Quantitative Economics*, 5(3):583–630. 2

Kotsadam, A. and Tolonen, A. K. (2015). African mining, gender, and local employment. *World Bank Policy Research Working Paper*, (7251). 76

Kremer, M., Miguel, E., and Thornton, R. (2009). Incentives to learn. *The Review of Economics and Statistics*, 91(3):437–456. 49

Laschever, R. (2009). The doughboys network: social interactions and labor market outcomes of world war i veterans. *Unpublished manuscript, Northwestern University*. 50

Lee, L.-f. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, 140(2):333–374. 50

Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics*, 28(4):825–860. 50

Loyalka, P., Song, Y., Wei, J., Zhong, W., and Rozelle, S. (2013). Information, college decisions and financial aid: Evidence from a cluster-randomized controlled trial in china. *Economics of Education Review*, 36:26–40. 2, 5

Luppino, M. and Sander, R. (2015). College major peer effects and attrition from the sciences. *IZA Journal of Labor Economics*, 4(1):1–23. 46, 52

Manski, C. (1992). Income and higher education. *Focus*, 14(3):14–19. 1

Manski, C. F. (1993a). Adolescent econometricians: How do youth infer the returns to schooling? In *Studies of supply and demand in higher education*, pages 43–60. University of Chicago Press. 1

Manski, C. F. (1993b). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542. 49

Martınez, L. R. (2016). Sources of revenue and government performance: Evidence from colombia. *Working Paper*. 100

Massé, F. and Camargo, J. (2012). Actores armados ilegales y sector extractivo en colombia. 74, 80

McGuigan, M., McNally, S., and Wyness, G. (2014). Student awareness of costs and benefits of educational decisions: Effects of an information campaign and media exposure. Technical report, Department of Quantitative Social Science-Institute of Education, University of London. 2, 3, 20, 25

McMahon, W. (2009). *Higher Learning, Greater Good: The Private and Social Benefits of Higher Education*. Johns Hopkins University Press. 1

Melguizo, T., Sanchez, F., and Velasco, T. (2016). Credit for Low-Income Students and Access to and Academic Performance in Higher Education in Colombia: A Regression Discontinuity Approach. *World Development*, 80:61–77. 8, 31

Mora, T. and Oreopoulos, P. (2011). Peer effects on high school aspirations: Evidence from a sample of close and not-so-close friends. *Economics of Education Review*, 30(4):575–581. 46, 47, 52

Nguyen, T. (2008). Information, role models and perceived returns to education: Experimental evidence from Madagascar. *Unpublished manuscript*. 2, 12

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101. 26

Oreopoulos, P. and Dunn, R. (2013). Information and College Access: Evidence from a Randomized Field Experiment. *Scandinavian Journal of Economics*, 115(1):3–26. 5, 25

Oreopoulos, P. and Petronijevic, U. (2013). Making college worth it: A review of research on the returns to higher education. Technical report, National Bureau of Economic Research. 2

Papay, J. P., Murnane, R. J., and Willett, J. B. (2015). The impact of test-score labels on human-capital investment decisions. *Journal of Human Resources*. 45

Pekkala-Kerr, S., Pekkarinen, T., Sarvimaki, M., and Uusitalo, R. (2015). Post-Secondary Education and Information on Labor Market Prospects: A Randomized Field Experiment. IZA Discussion Papers 9372, Institute for the Study of Labor (IZA). 3, 5, 20, 25

Perry, G. and Olivera, M. (2009). El impacto del petróleo y la minería en el desarrollo regional y local en colombia. 100

Portafolio (2011). Denuncian evasión millonaria en pago de regalías del oro. http://www.portafolio.co/economia/denuncian-evasion-millonaria-pago-regalias-del-oro. Accessed: 2016-01-20. 80

Ray, D. (2006). Aspirations, poverty, and economic change. *Understanding poverty*, pages 409–21. 26

Reboredo, J. C. (2013). Is gold a safe haven or a hedge for the us dollar? implications for risk management. *Journal of Banking & Finance*, 37(8):2665–2676. 78

Rettberg, A. and Ortiz-Riomalo, J. F. (2014). Golden conflict: Exploring the relationship between gold mining, armed conflict, and criminality in colombia. 74, 80

Reyes, L., Rodríguez, J., and Urzúa, S. S. (2016). Heterogeneous economic returns to postsecondary degrees: Evidence from chile. *Journal of Human Resources*, 51(2):416–460. 45

Romero, M. and Saavedra, S. (2015). The effect of gold mining on the health of newborns. *Mimeo*. 73, 80

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics*, 116(2):681–704. 46, 48, 50, 51

Sachs, J. D. and Warner, A. M. (1995). Natural resource abundance and economic growth. Technical report, National Bureau of Economic Research. 76

Santos, R. J. (2014). Not all that glitters is gold: Gold boom, child labor and schooling in colombia. *Documento CEDE*, (2014-31). 77

Solis, A. (2013). Credit Access and College Enrollment. Working Paper Series 2013:12, Uppsala University, Department of Economics. 1

Stijns, J.-P. (2006). Natural resource abundance and human capital accumulation. *World Development*, 34(6):1060–1083. 77

The Economist (2013). Digging itself out of a hole. http://www.economist.com/news/business/21599011-government-struggles-contain-public-backlash-against-miners-digging-itself-out. Accessed: 2016-01-12. 78

The Telegraph (2016). Colombia investigates gold trades for suspected cocaine money-laundering. http://www.telegraph.co.uk/finance/financial-crime/10857780/Colombia-investigates-gold-trades-for-suspected-cocaine-money-laundering.html. Accessed: 2015-12-11. 80

UNEP, U. (2015). Mineral yearbook. 76

UNODC, U. (2015). Colombia survey 2014. 80, 88

USGS, U. (2013). Mineral yearbook. `http://minerals.usgs.gov/minerals/pubs/commodity/gold/`. Accessed: 2015-06-10. 77

Wilson, N. (2012). Economic booms and risky sexual behavior: evidence from zambian copper mining cities. *Journal of Health Economics*, 31(6):797–812. 76

Wiswall, M. and Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, 82(2):791–824. 25, 45

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources*, 48(3):545–595. 45

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics*, 85(1):9–23. 50

Appendix A

Additional Tables and Figures of Chapter 1

Table A.1: Attrition diagnostics

	BHELPS: Baseline to Follow-Up	BHELPS (baseline) to ICFES	BHELPS (baseline) to ICFES-SNIES
	(1)	(2)	(3)
Panel A: Attrition Rates			
Baseline N	6,636	6,636	6,636
Final N	6,141	6,323	6,303
Attrition Rate	0.075	0.047	0.050
Panel B: Random attrition tests (OLS)			
Treatment	-0.009	-0.021	-0.021
	(0.039)	(0.016)	(0.016)
R^2	0.000	0.002	0.002

Source: Authors' calculations from ICFES, SNIES, and BHELPS survey.
Notes: * Significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors are clustered at school-level.

Table A.2: Treatment Effects on Beliefs: Robustness

	Robust Premium Error Degree and Field			Robust Premium Error Degree, College, and Field		
	All (1)	Under (2)	Over (3)	All (1)	Under (2)	Over (3)
Panel A: After, Matched with baseline						
Treat	-0.035	0.194	-0.066	0.027	-0.048	0.045
	(0.091)	(0.200)	(0.091)	(0.219)	(0.511)	(0.233)
Observations	4,011	802	3,209	2,811	596	2,215
Panel B: Difference-in-differences						
Treat × Post	0.146	0.106	0.122	-0.003	-0.005	-0.022
	(0.099)	(0.265)	(0.098)	(0.100)	(0.316)	(0.099)
Post	-0.091	1.243***	-0.337***	-0.116	0.977***	-0.347***
	(0.064)	(0.160)	(0.068)	(0.072)	(0.170)	(0.071)
Observations	8,020	1,468	6,552	5,618	1,114	4,504

Source: Authors' calculations from BHELPS survey.
Notes: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each column and panel correspond to a separate OLS regression. Panels A controls for student and household attributes (male, age, age squared, family income, and parental education) and school characteristics (average SABER 11 score in previous years, has computer lab, shift indicators, and school size). Panel B presents coefficients for difference-in-difference regressions that control for individual fixed-effects. Standard errors are clustered at school-level.

Table A.3: Treatment Effects by Family Income and Gender (balanced sample)

	Knows ICETEX (1)	Overall score (2)	Math (3)	Language (4)	College Enrollment (5)	Private College (6)	Top-10 College (7)	Academic Degree (8)	STEM Degree (9)
Panel A: Treatment effects by family income									
Low income (<=2 MW)	0.061***	-0.028	0.017	-0.050	-0.009	0.021**	0.000	0.007	0.007
	(0.018)	(0.038)	(0.040)	(0.038)	(0.023)	(0.011)	(0.002)	(0.009)	(0.006)
Middle income (>2 MWs)	0.028*	0.048	0.093*	0.073	0.031	0.002	0.017***	0.019	0.013
	(0.015)	(0.049)	(0.049)	(0.047)	(0.027)	(0.021)	(0.005)	(0.017)	(0.013)
P-value (Low=Middle)	0.080	0.165	0.108	0.025	0.137	0.382	0.005	0.507	0.706
Observations	5,427	5,414	5,414	5,414	5,401	5,401	5,401	5,401	5,401
Panel B: Treatment effects by Gender									
Female	0.029	-0.027	0.025	-0.040	-0.021	0.004	0.004	0.006	0.002
	(0.018)	(0.039)	(0.042)	(0.043)	(0.026)	(0.015)	(0.003)	(0.011)	(0.007)
Male	0.069***	0.030	0.065	0.036	0.033	0.021	0.008*	0.016	0.016
	(0.019)	(0.044)	(0.048)	(0.039)	(0.022)	(0.014)	(0.005)	(0.013)	(0.010)
P-value (Female=Male)	0.069	0.236	0.411	0.149	0.051	0.348	0.513	0.584	0.290
Observations	5,427	5,414	5,414	5,414	5,401	5,401	5,401	5,401	5,401

Source: Authors' calculations from ICFES, SNIES, and BHELPS survey.
Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each column and panel corresponds to a separate OLS regression that controls for student and household attributes (male, age, age squared, family income, and parental education) and school characteristics (average SABER 11 score in previous years, has computer lab, shift indicators, and school size). Standard errors are clustered at school-level.

Table A.4: Treatment Effects by Direction of Belief Error (baseline matched to administrative data)

	Knows ICETEX (1)	Overall score (2)	Math (3)	Language (4)	College Enrollment (5)	Private College (6)	Top-10 College (7)	Academic Degree (8)	STEM Degree (9)
Panel A: Average effects									
Treat	0.039***	-0.012	0.032	-0.010	0.012	0.015	0.005*	0.009	0.008
	(0.014)	(0.030)	(0.034)	(0.029)	(0.019)	(0.010)	(0.002)	(0.009)	(0.006)
Panel B: Treatment effects by direction of belief error									
Underestimates	0.053	-0.003	0.102	0.002	0.038	0.002	0.000	0.022	0.023*
	(0.034)	(0.076)	(0.075)	(0.077)	(0.031)	(0.023)	(0.005)	(0.020)	(0.013)
Overestimates	0.050***	-0.019	0.019	-0.014	0.006	0.018	0.005*	0.006	0.007
	(0.016)	(0.031)	(0.034)	(0.032)	(0.019)	(0.011)	(0.003)	(0.009)	(0.007)
P-value (low=middle)	0.933	0.845	0.266	0.845	0.318	0.505	0.450	0.423	0.285
Observations	6,003	6,309	6,309	6,309	6,289	6,289	6,289	6,289	6,289

Source: Authors' calculations from ICFES, SNIES, and BHELPS survey.
Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each column and panel corresponds to a separate OLS regression that controls for student and household attributes (male, age, age squared, family income, and parental education) and school characteristics (average SABER11 score in previous years, has computer lab, shift indicators, and school size). Standard errors are clustered at school-level.

Table A.5: Treatment Effects by Non-Cognitive Factors (baseline matched to administrative data)

	Knows ICETEX (1)	Overall score (2)	Math (3)	Language (4)	College Enrollment (5)	Private College (6)	Top-10 College (7)	Academic Degree (8)	STEM Degree (9)
Panel A: Treatment effects by risk aversion									
Risk loving	0.038	0.042	0.093	0.049	0.036	0.022	0.014	0.029	0.035***
	(0.030)	(0.079)	(0.083)	(0.071)	(0.038)	(0.025)	(0.009)	(0.018)	(0.013)
Risk averse	0.052***	-0.024	0.020	-0.023	0.006	0.013	0.003	0.006	0.005
	(0.016)	(0.033)	(0.035)	(0.032)	(0.019)	(0.012)	(0.002)	(0.009)	(0.007)
P-value (loving=averse)	0.653	0.440	0.381	0.360	0.414	0.742	0.227	0.214	0.024
Observations	5,225	6,076	6,076	6,076	6,057	6,057	6,057	6,057	6,057
Panel B: Treatment effects by self-concept									
Low	0.064***	-0.000	0.047	-0.020	0.013	0.018	0.004	0.007	0.006
	(0.017)	(0.036)	(0.042)	(0.037)	(0.022)	(0.012)	(0.002)	(0.009)	(0.005)
High	0.027	0.011	0.044	0.037	0.018	0.013	0.007	0.015	0.013
	(0.019)	(0.043)	(0.045)	(0.040)	(0.025)	(0.017)	(0.005)	(0.014)	(0.011)
P-value (Low=High)	0.092	0.812	0.941	0.260	0.848	0.769	0.560	0.560	0.531
Observations	5,382	6,259	6,259	6,259	6,239	6,239	6,239	6,239	6,239
Panel C: Treatment effects by self-efficacy									
Low	0.043**	-0.046	0.016	-0.058	0.007	0.012	0.003	0.001	0.002
	(0.017)	(0.037)	(0.042)	(0.035)	(0.020)	(0.011)	(0.003)	(0.011)	(0.007)
High	0.052***	0.069	0.081*	0.091*	0.019	0.020	0.007*	0.024*	0.019**
	(0.019)	(0.045)	(0.047)	(0.046)	(0.024)	(0.015)	(0.004)	(0.012)	(0.009)
P-value (Low=High)	0.650	0.028	0.260	0.007	0.629	0.608	0.424	0.150	0.119
Observations	5,378	6,248	6,248	6,248	6,228	6,228	6,228	6,228	6,228
Panel D: Treatment effects by perceived likelihood of enrollment									
Low	0.101***	-0.031	-0.004	-0.032	0.013	-0.004	-0.002	0.003	-0.001
	(0.034)	(0.050)	(0.053)	(0.056)	(0.032)	(0.015)	(0.002)	(0.012)	(0.008)
High	0.038***	-0.004	0.042	-0.006	0.011	0.017	0.006**	0.013	0.012
	(0.014)	(0.034)	(0.037)	(0.033)	(0.019)	(0.012)	(0.003)	(0.010)	(0.007)
P-value (Low=High)	0.041	0.649	0.470	0.699	0.937	0.243	0.017	0.512	0.206
Observations	5,169	6,014	6,014	6,014	5,995	5,995	5,995	5,995	5,995

Source: Authors' calculations from ICFES, SNIES, and BHELPS survey.
Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each column and panel corresponds to a separate OLS regression that controls for student and household attributes (male, age, age squared, family income, and parental education) and school characteristics (average SABER 11 score in previous years, has computer lab, shift indicators, and school size). Standard errors are clustered at school-level.

Table A.6: Baseline Balance in Student Aspirations

	Control		Treatment		Difference
	Mean	(SD)	Mean	(SD)	P-value
Enroll	0.988	(0.108)	0.988	(0.108)	1.000
Public College	0.628	(0.484)	0.629	(0.483)	0.944
Private College	0.220	(0.415)	0.234	(0.423)	0.403
Top-10 College	0.451	(0.498)	0.470	(0.499)	0.442
Academic degree (4-year)	0.886	(0.317)	0.897	(0.304)	0.359
Vocational degree (2-year)	0.087	(0.281)	0.081	(0.273)	0.554
STEM degree	0.403	(0.491)	0.430	(0.495)	0.089

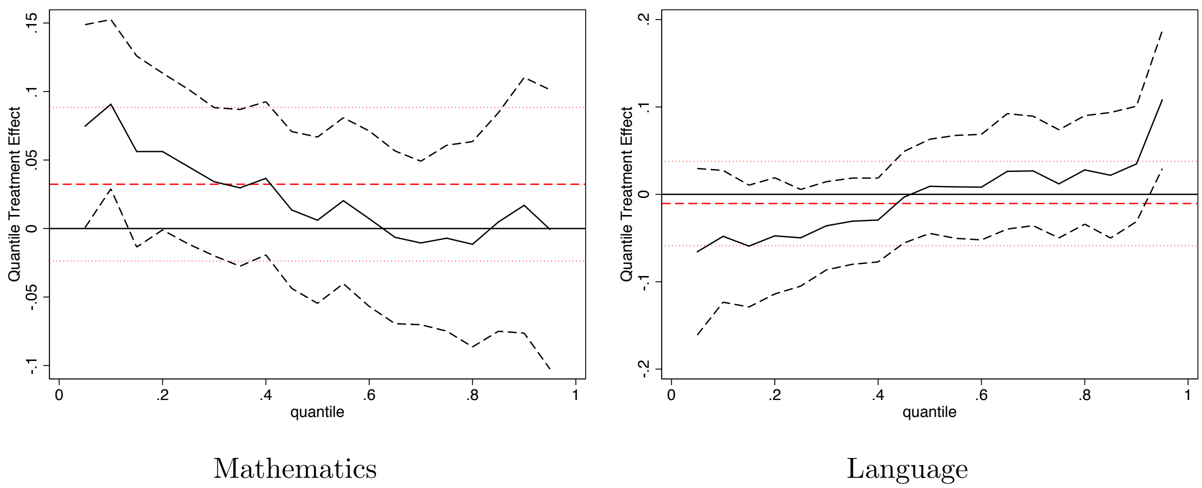
Source: Authors' calculations from BHELPS survey on balanced sample.
Notes: The last two columns present the difference in means and p-values between treatment and control groups calculated by regression with clustered standard errors at the school-level.

Table A.7: Treatment Effects on College Aspirations

	Enroll College (1)	Private College (2)	Top-10 College (3)	Academic Degree (4)	STEM Degree (5)
Panel A: After, All students in follow-up					
Treat	0.001	0.016	0.020	0.021*	0.033**
	(0.003)	(0.015)	(0.018)	(0.012)	(0.013)
Observations	6,072	6,072	6,072	6,072	6,072
Panel B: After, Matched with baseline					
Treat	0.002	0.019	0.020	0.021*	0.036**
	(0.003)	(0.015)	(0.019)	(0.013)	(0.014)
Observations	5,485	5,485	5,485	5,485	5,485
Panel C: ANOVA					
Treat	0.001	0.014	0.012	0.017	0.020*
	(0.003)	(0.014)	(0.017)	(0.012)	(0.012)
Observations	5,485	5,485	5,485	5,485	5,485
Panel D: Difference-in-differences					
Treat × Post	-0.001	0.003	-0.004	0.004	0.007
	(0.004)	(0.016)	(0.023)	(0.013)	(0.014)
Post	0.004	0.009	0.000	-0.027***	-0.006
	(0.003)	(0.012)	(0.018)	(0.009)	(0.010)
Observations	11,006	11,006	11,006	11,006	11,006
Mean(y) at baseline	0.983	0.228	0.449	0.877	0.410

Source: Authors' calculations from BHELPS survey.
Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each column and panel correspond to a separate OLS regression. Panels A and B control for student and household attributes (male, age, age squared, family income, and parental education) and school characteristics (average SABER 11 score in previous years, has computer lab, shift indicators, and school size). Panel C presents coefficients for difference-in-difference regression that control for individual fixed-effects. Standard errors are clustered at school-level.

Figure A.1: Quantile Treatment Effects for SABER 11 Test Scores



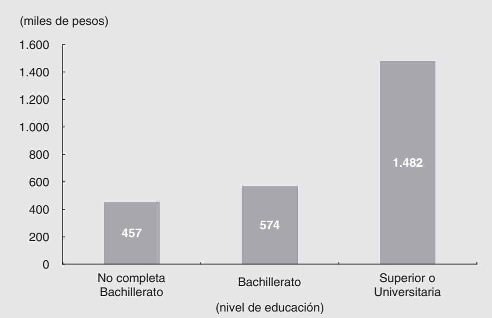
Source: Authors' elaboration from ICFES and BHELPS survey.
Notes: Estimates based on baseline matched to administrative data (N=6,309). 90% Confidence intervals in black dashed/red dotted lines. OLS estimate in red dashed line. Standard errors clustered at the school-level.

Figure A.2: Student Handout: Original Version

¡La educación superior paga!

La relación entre estudios e ingresos

La educación superior es un factor determinante de la situación económica y por tanto la calidad de vida de las familias. En el siguiente gráfico se presentan los salarios promedio por nivel educativo en Bogotá.



Fuente: Encuesta de Hogares 2011, DANE

Como se puede observar, mayor educación se traduce en salarios más altos. Sólo con terminar el Bachillerato se pasa de ganar 457.000 a 574.000 por mes. El salto es más evidente para aquellos con un título de nivel superior, ya que el salario promedio mensual crece a 1.482.000. Estas estadísticas presentan un mensaje claro: vale la pena estudiar.

¿Cómo puedo averiguar cuanto ganaría en la carrera que a mí me interesa?

Es probable que usted ya tenga una idea sobre las carreras que le interesarían y la institución donde quisiera realizar estos estudios. Si es así, ¿hay alguna manera de saber cuánto puede esperar ganar en su situación específica?

Existen dos lugares donde pueden consultar el salario promedio de los graduados por institución y carreras. Estas son:

1. Calculadora de salarios promedios para graduados: www.finanzaspersonales.com.co

Esta página cuenta con una herramienta que le permite consultar el salario promedio por región, institución educativa, programa de estudio y género de las personas que obtuvieron su título entre 2001-2011.

¿Cómo funciona?

- Acceda al enlace y busque la *Calculadora de Salario por profesión para Graduados*
- Escoja la región donde quiere realizar la búsqueda (por ejemplo, Bogotá)
- Seleccione la institución donde quiere realizar sus estudios y el programa que planea cursar

2. Observatorio laboral del Ministerio de Educación: www.graduadoscolombia.edu.co

Está página también provee información sobre los salarios promedios de personas con título de educación superior para toda Colombia. Además, le permite conocer las perspectivas laborales del programa de estudio de su interés.

¿Cómo funciona?

- Acceda al enlace y busque el botón rojo que dice *Sistema de información del Observatorio Laboral*.
- Si quiere conocer el número de graduados por carrera, acceda a la pestaña que dice "Perfil nacional". Después, escoja el departamento donde planea estudiar y obtendrá los datos de graduados por área de estudio.

Si desea saber cuántos individuos en su área de interés tienen un empleo formal (cotizando a la seguridad social) y cuanto ganan en promedio vaya a ¨Vinculación laboral recién graduados¨. Aquí tiene la opción de buscar por institución o por carrera.

Recuerde que estas páginas le permiten conocer el salario promedio de los profesionales graduados en su área de interés.

¿Qué necesito para entrar a la Universidad y la carrera que me interesa?

1. Buenos resultados académicos: Uno de los criterios más importantes a la hora de buscar admisión a una institución de educación superior es el rendimiento académico. Muchas instituciones utilizan el puntaje del ICFES (SABER 11), y otras instituciones como la Universidad Nacional que tienen su propio examen de admisión. En cualquier caso, estudiar aumenta las posibilidades de ser admitido y también las posibilidades de acceder a becas o financiación.

2. Financiación: Existen varias maneras de financiar la educación superior en Colombia. En general, tendrán preferencia los alumnos de escasos recursos y buen desempeño académico. Las siguientes son algunas opciones a tener en cuenta:

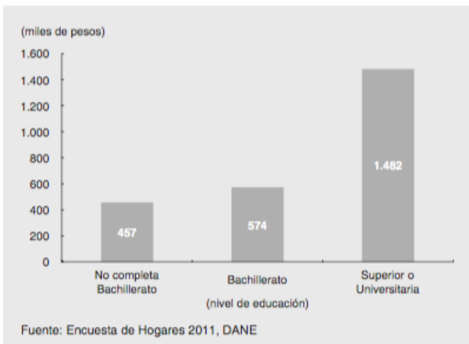
- Becas proveídas por cada institución por mérito académico y/o escasos recursos. Consulte las políticas de beca ya que estas son diferentes para cada institución.
- ICETEX: http://www.icetex.gov.co
- Secretaría de Educación de Bogotá (Banco de cupos, Fondo de Financiamiento de Educación Superior de Bogotá): http://www.sedbogota.edu.co/index.php/educacion-superior.html

Figure A.3: Student Handout: English Translation

Post-secondary education pays!

The relation between studies and income

Higher education is a determining factor of wages and the quality of life of families. The following figure presents average wages by level of completed education in Bogotá:



Clearly, more education is related with higher wages. By only finishing high school, wages move from 457,000 to 574,000 pesos each month. The difference is even more marked for those with a college degree, since their average monthly wage increases to 1,492,000. These statistics present a clear pattern: studying is worth it.

How can I learn about how much people earn who finished the degree I'm interested in?

It is very likely that you already have a good idea about the degrees and institutions where you would like to pursue your studies. If this is true, is there a way to know how much I could expect to earn?

There are two places where you can obtain information on average wages for graduates by institution and degree. These are:

1. Average wage calculator for graduates: www.finanzaspersonales.com.co

This website counts with a tool that allows to calculate average wages by region, institution, degree and gender of people who graduated between 2001 and 2011.

How does it work?

- Visit the website and search for *Wage calculator by degree for Graduates.*
- Select the region where you are interested in searching (e.g. Bogotá)
- Select the institution and the degree you are interested in evaluating

2. Labor Observatory of the Ministry of Education: www.graduadoscolombia.edu.co

This website also provides information about average wages for the whole country. Additionally, you can learn about the labor prospects for your degree of interest

How does it work?

- Visit the website and click on the red button reading *Information System of the Labor Observatory*
- If you would like to know the number of graduates by degree, click on the "National Profile" tab. Next, select the department where you plan to study and you will find data on graduates by degree.

If you are interested in the number of individuals who pursued your degree of interest who have a formal job (paying social security) and how much they earn on average, select "*labor link of recent graduates*". Here you have the option to search by institution and degree.

Remember that these websites allow to learn about the average wages of recent graduates for your degree of interest.

What will I need to enroll in a University and in my degree of interest?

1. **Good academic results**: One of the main criteria for admissions in University if academic performance. Many institutions use the ICFES (SABER 11) score, and other institutions like the National University also have their own admissions test. Nevertheless, studying will increase the probability of being admitted and also of obtaining financial aid or financing.
2. **Financing:** There are many ways to finance higher education in Colombia. In general, financing institutions have preferences for students of low income and good academic performance. The following are some organizations to keep in mind:

- Scholarships provided by each institution according to academic merit of financial need. Consult the scholarship policies for each institution given that they may differ.
- ICETEX: http://www.icetex.gov.co
- Secretary of Education in Bogotá (FDFESBO): http://www.sedbogota.edu.co/index.php/ed- ucacion-superior.html

Appendix B

Additional Tables and Figures of Chapter 3

Table B.1: Effect of Gold Mining on School Enrollment and Progress Throughout the Year (20 km Neighborhood): First-stage Estimates

	Primary (1)	Middle (2)	High (3)
Active titles	0.587***	0.546***	0.519***
	(0.024)	(0.073)	(0.084)
First-stage F	600.8	55.5	37.9
Mining deforestation	0.252***	0.291***	0.315***
	(0.010)	(0.031)	(0.042)
First-stage F	666.0	90.6	56.4
Observations	47,740	10,509	6,214

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate instrumental variable first-stage regression that controls for school and year fixed effects, and municipal-specific time trends. Mining intensity measures are instrumented with the interaction between gold deposits in the neighborhood and international prices. Standard errors are clustered at school level. The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2004-2012. A mine is considered in the neighborhood of a school if the distance is smaller or equal to 20 km. Active titles and mining deforestation, as defined in Section 3.3.2, and the instrument are normalized with mean zero and standard deviation equal to one.

Table B.2: Effect of Gold Mining on School Enrollment in the Following Year (20 km Neighborhood)

	Active titles			Mining deforestation		
	Primary	Middle	High	Primary	Middle	High
	(1)	(2)	(3)	(4)	(5)	(6)
OLS	0.707	1.056	2.556	-0.414	0.315	-0.117
	(0.458)	(1.327)	(0.972)	(0.487)	(1.567)	(0.987)
IV	-1.155	-0.472	8.426	-2.386	-0.698	11.385
	(5.103)	(17.497)	(11.116)	(4.305)	(13.278)	(9.033)
mean(y)	75.554	191.036	98.68	75.554	191.036	98.68
First-stage F	589.6	54.3	36.7	709.5	153.2	88.5
Observations	42,646	9,738	5,775	42,602	9,588	5,703

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate panel regression that controls for school and year fixed effects, and municipal-specific time trends. IV regressions instrument the mining intensity measures with the interaction between gold deposits in the neighborhood and international prices. Standard errors are clustered at school level. The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2004-2012. A mine is considered in the neighborhood of a school if the distance is smaller or equal to 20 km. Active titles and mining deforestation, as defined in Section 3.3.2, are normalized with mean zero and standard deviation equal to one.

Table B.3: Effect of Gold Mining on Students Taking the Exit Exam (20 km Neighborhood): First-stage Estimates

	Students per school	Exam score	Student works*
	(1)	(2)	(3)
Active titles	0.452***	1.074***	1.144***
	(0.074)	(0.064)	(0.043)
First-stage F	37.5	220.1	554.1
Mining deforestation	0.265***	1.302***	1.093***
	(0.034)	(0.101)	(0.055)
First-stage F	61.9	120.2	308.4
Observations	5,180	318,226	195,879

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate instrumental variable first-stage regression that controls for school and year fixed effects, and municipal-specific time trends. Mining intensity measures are instrumented with the interaction between gold deposits in the neighborhood and international prices. Individual regressions also control for the students' age and gender. Standard errors are clustered at school level. The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2004-2012, except for work situation that is restricted to 2008-2012. A mine is considered in the neighborhood of a school if the distance is smaller or equal to 20 km. Active titles and mining deforestation,as defined in Section 3.3.2, the instrument, and test scores are normalized with mean zero and standard deviation equal to one.

Table B.4: Effect of Gold Mining on School Attendance and Child Labor (20 km Neighborhood): First-stage Estimates

	All (1)	6-8 (2)	9-11 (3)	12-14 (4)	15-17 (5)
Active titles	0.802*** (0.091)	0.860*** (0.112)	0.730*** (0.072)	0.870*** (0.118)	0.753*** (0.105)
First-stage F	76.9	59.2	101.5	54.3	51.7
Mining deforestation	0.370*** (0.022)	0.361*** (0.033)	0.384*** (0.023)	0.369*** (0.027)	0.379*** (0.022)
First-stage F	274.0	117.3	268.6	180.2	284.2
Observations	3,627	846	868	963	950

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate instrumental variable first-stage regression that controls for municipal and year fixed effects, and students characteristics (age, gender, household sized and parents' education). Mining intensity measures are instrumented with the interaction between gold deposits in the neighborhood and international prices. Standard errors are clustered at DHS cluster level. The sample includes all urban households in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2005-2010. A mine is considered in the neighborhood of a cluster if the distance is smaller or equal to 20 km. Active titles and mining deforestation, as defined in Section 3.3.2, and the instrument are normalized with mean zero and standard deviation equal to one.

Table B.5: Effect of Gold Mining on School Attendance and Child Labor: Urban and Rural Households (20 km Neighborhood)

	Active titles					Mining deforestation				
	All	6-8	9-11	12-14	15-17	All	6-8	9-11	12-14	15-17
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. School attendance										
OLS	-0.007	0.012	-0.026	0.044	-0.008	0.063**	-0.039	0.057	0.098**	0.033
	(0.027)	(0.041)	(0.023)	(0.037)	(0.058)	(0.032)	(0.062)	(0.037)	(0.044)	(0.058)
IV	0.003	0.004	-0.006	0.055	-0.028	0.005	0.008	-0.012	0.107	-0.053
	(0.036)	(0.047)	(0.026)	(0.056)	(0.079)	(0.069)	(0.093)	(0.050)	(0.109)	(0.151)
Mean(y)	0.795	0.613	0.955	0.906	0.694	0.795	0.613	0.955	0.906	0.694
Panel B. Works										
OLS	0.018	-0.010*	0.054**	0.015	-0.027	-0.008	-0.001	0.045*	-0.103***	0.067
	(0.015)	(0.006)	(0.023)	(0.026)	(0.053)	(0.019)	(0.008)	(0.026)	(0.040)	(0.055)
IV	0.054**	0.001	0.082**	0.033	0.073	0.105**	0.002	0.154**	0.064	0.138
	(0.025)	(0.007)	(0.034)	(0.035)	(0.067)	(0.050)	(0.015)	(0.068)	(0.069)	(0.127)
Mean(y)	0.106	0.011	0.045	0.114	0.248	0.106	0.011	0.045	0.114	0.248
First-stage F	508.9	568.3	461.3	219.1	500.5	293.2	226.0	320.7	214.9	237.9
Observations	7,058	1,677	1,781	1,815	1,785	7,058	1,677	1,781	1,815	1,785

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate regression that controls for municipal and year fixed effects, and students characteristics (age, gender, household sized and parents' education). IV regressions instrument the mining intensity measures with the interaction between gold deposits in the neighborhood and international prices. Standard errors are clustered at DHS cluster level. The sample includes all urban and rural households in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2005-2010. A mine is considered in the neighborhood of a cluster if the distance is smaller or equal to 20 km. Active titles and mining deforestation, as defined in Section 3.3.2, are normalized with mean zero and standard deviation equal to one.

Table B.6: Effect of Gold Mining on School Enrollment and Progress Throughout the Year
(IV only, 20 km Neighborhood): Controlling for Homicide and Royalties

	Active titles			Mining deforestation		
	Primary	Middle	High	Primary	Middle	High
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Enrollment (beginning of the year)						
Mining	-2.359**	-1.582	-4.844*	-5.461**	-2.94	-7.933*
	(0.940)	(2.290)	(2.632)	(2.193)	(4.152)	(4.252)
Homicide rate	-0.003	0.002	-0.010*	-0.001	0.003	-0.007
	(0.003)	(0.005)	(0.005)	(0.003)	(0.005)	(0.006)
Royalties	0.142	-5.704*	4.739	1.238	-5.067	5.720*
	(1.905)	(3.450)	(3.497)	(1.950)	(3.571)	(3.354)
Mean(y)	78.249	81.009	87.39	78.249	81.009	87.39
Panel B. Grade promotion rate						
Mining	-2.359**	-1.582	-4.844*	-5.461**	-2.94	-7.933*
	(0.940)	(2.290)	(2.632)	(2.193)	(4.152)	(4.252)
Homicide rate	-0.003	0.002	-0.010*	-0.001	0.003	-0.007
	(0.003)	(0.005)	(0.005)	(0.003)	(0.005)	(0.006)
Royalties	0.142	-5.704*	4.739	1.238	-5.067	5.720*
	(1.905)	(3.450)	(3.497)	(1.950)	(3.571)	(3.354)
Mean(y)	78.249	81.009	87.39	78.249	81.009	87.39
Panel C. Grade repetition rate						
Mining	0.314	0.837	2.948*	0.727	1.556	4.828*
	(0.634)	(1.403)	(1.596)	(1.468)	(2.564)	(2.627)
Homicide rate	0.004**	-0.004	0.000	0.004*	-0.005	-0.002
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Royalties	-0.532	0.486	0.338	-0.678	0.149	-0.259
	(1.333)	(1.618)	(1.296)	(1.356)	(1.714)	(1.465)
Mean(y)	9.373	6.540	4.499	9.373	6.540	4.499
Panel C. Dropout rate						
Mining	1.263**	1.586	3.227**	2.924**	2.948	5.284**
	(0.509)	(1.555)	(1.612)	(1.187)	(2.846)	(2.555)
Homicide rate	0.001	0.003	0.005	0.000	0.001	0.003
	(0.002)	(0.003)	(0.003)	(0.002)	(0.004)	(0.004)
Royalties	0.554	3.408	-4.429	-0.033	2.769	-5.082*
	(1.079)	(2.252)	(3.075)	(1.109)	(2.329)	(2.891)
Mean(y)	6.484	7.886	4.878	6.484	7.886	4.878
First-stage F	599.703	55.255	37.89	669.495	92.325	57.3
Observations	47,688	10,497	6,206	47,688	10,497	6,206

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate panel regression that controls for municipal annual homicide rate and per capita royalties, school and year fixed effects, and municipal-specific time trends. Standard errors are clustered at school level. IV regressions instrument the mining intensity measures with the interaction between gold deposits in the neighborhood and international prices. The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2004-2012. A mine is considered in the neighborhood of a school if the distance is smaller or equal to 20 km. Active titles and mining deforestation, as defined in Section 3.3.2, are normalized with mean zero and standard deviation equal to one. Homicide rates are expressed in terms of homicides per 100,000 residents, and royalties in millions of Colombian pesos per capita.

Table B.7: Effect of Gold Mining on Students Taking the Exit Exam (IV only, 20 km Neighborhood): Controlling for Homicide and Royalties

	Active titles			Mining deforestation		
	Students per school (1)	Exam score (2)	Student works* (3)	Students per school (4)	Exam Score (5)	Student works* (6)
Mining	7.879*	0.013	0.001	13.494*	0.011	0.001
	(4.427)	(0.009)	(0.003)	(7.448)	(0.007)	(0.003)
Homicide rate	0.019*	0.000	0.000	0.011	0.000	0.000
	(0.011)	(0.000)	(0.000)	(0.011)	(0.000)	(0.000)
Royalties	-1.066	0.013	0.039*	-3.416	0.008	0.039*
	(6.378)	(0.036)	(0.020)	(5.883)	(0.036)	(0.020)
Mean(y)	50.83	0.000	0.129	62.163	0.000	0.129
First-stage F	37.338	282.539	728.542	50.83	164.226	396.825
Observations	5,174	318,028	195,729	5,174	318,028	195,729

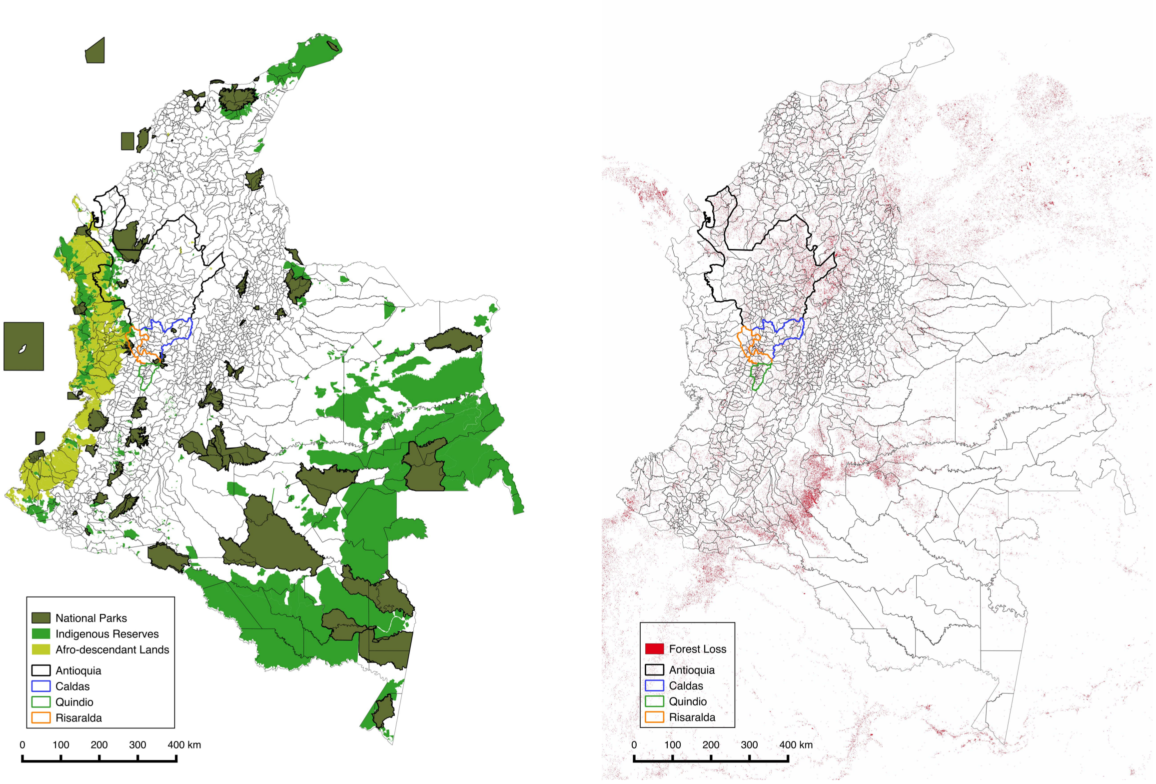
Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate regression that controls for municipal annual homicide rate and per capita royalties, school and year fixed effects, and municipal-specific time trends. Individual regressions also control for the students' age and gender. IV regressions instrument the mining intensity measures with the interaction between gold deposits in the neighborhood and international prices. Standard errors are clustered at school level. The sample includes all geocoded schools in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2004-2012, except for work situation that is restricted to 2008-2012. A mine is considered in the neighborhood of a school if the distance is smaller or equal to 20 km. Active titles and mining deforestation, as defined in Section 3.3.2, and test scores are normalized with mean zero and standard deviation equal to one. Homicide rates are expressed in terms of homicides per 100,000 residents, and royalties in millions of Colombian pesos per capita.

Table B.8: Effect of Gold Mining on School Attendance and Child Labor (IV only, 20 km Neighborhood): Controlling for Homicide and Royalties

	Active titles					Mining deforestation				
	All	6-8	9-11	12-14	15-17	All	6-8	9-11	12-14	15-17
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. School attendance										
Mining	-0.02	-0.01	-0.050***	-0.029	-0.08	-0.043	-0.023	-0.098**	-0.063	-0.164
	(0.038)	(0.080)	(0.017)	(0.041)	(0.077)	(0.082)	(0.190)	(0.039)	(0.093)	(0.165)
Homicide rate	0.001	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000
	(0.001)	(0.001)	0.000	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Royalties	2.798	-1.217	-0.08	-4.629**	0.758	4.252	0.277	-0.663	-5.419**	-0.146
	(4.371)	(4.486)	(2.181)	(2.019)	(4.187)	(4.486)	(4.612)	(2.336)	(2.591)	(4.161)
Mean(y)	0.838	0.647	0.977	0.938	0.779	0.838	0.647	0.977	0.938	0.779
Panel B. Works										
Mining	0.047*	0.002	0.073***	-0.007	0.107	0.101*	0.004	0.144**	-0.016	0.218
	(0.033)	(0.072)	(0.025)	(0.014)	(0.045)	(0.069)	(0.106)	(0.054)	(0.039)	(0.096)
Homicide rate	0.000	0.002	0.001	0.000	0.001	0.000	0.002	0.001*	0.001*	0.001
	(0.001)	(0.001)	0.000	0.000	(0.001)	(0.001)	(0.002)	0.000	0.000	(0.001)
Royalties	0.605	1.231	0.321	0.222	0.997	1.340	2.279	1.710	1.618	2.222
	(2.440)	(6.075)	(1.921)	(0.871)	(3.327)	(2.768)	(6.336)	(2.008)	(1.008)	(3.449)
Mean(y)	0.080	0.008	0.029	0.082	0.187	0.080	0.008	0.029	0.082	0.187
First-stage F	382.1	54.5	240.0	46.6	154.4	928.0	134.8	229.2	211.9	289.8
Observations	3,628	846	868	963	951	3,628	846	868	963	951

Note: * Significant at 10%; ** significant at 5%; *** significant at 1%. Each coefficient corresponds to a separate regression that controls for municipal annual homicide rate and per capita royalties, municipal and year fixed effects, and students characteristics (age, gender, household sized and parents' education). IV regressions instrument the mining intensity measures with the interaction between gold deposits in the neighborhood and international prices. Standard errors are clustered at DHS cluster level. The sample includes all urban households in non-metropolitan areas from Antioquia and the Coffee Region. The period of study is 2005-2010. A mine is considered in the neighborhood of a cluster if the distance is smaller or equal to 20 km. Active titles and mining deforestation, as defined in Section 3.3.2, are normalized with mean zero and standard deviation equal to one. Homicide rates are expressed in terms of homicides per 100,000 residents, and royalties in millions of Colombian pesos per capita.

Figure B.1: Protected Areas and Overall Deforestation



(a) Protected Areas   **(b)** Forest Loss (2001-2012)

Source: (a) SIGOT (b) Hansen et al. (2013).
Notes: (b) Each pixel corresponds to 30 m^2 of forest loss in the period 2001-2012.