

## SUBJECTIVE EARNINGS AND ACADEMIC EXPECTATIONS OF TERTIARY EDUCATION IN COLOMBIA

Luis Fernando Gamboa

Paul Andrés Rodríguez-Lesmes\*

---

Article Info: Received 21 July 2017; accepted 22 January 2018

---

### ABSTRACT

#### *JEL Classification*

I25

J24

D84

#### *Keywords:*

Subjective expectations

Colombian education

tertiary education

We analysed the academic expectations and earning expectations of a nationally representative sample of Colombian students who were finishing their upper-secondary education (high school) and considering three potential scenarios: whether to finish their studies, enrol in a vocational career, or pursue a college degree. We found that these students' earning expectations are correlated with local labour market wages. However, they expect earnings as university graduates that are significantly above the current observed earnings, which is not the case for the other two levels of education. We also found that earning and academic expectations (test scores and odds of attending college) are closely related to their family socioeconomic background, school and municipality characteristics, and even reflect aspects such as gender gaps or private school premiums. Finally, both academic and earnings expectations are related to actual realised test scores.

---

### INGRESOS SUBJETIVOS Y EXPECTATIVAS ACADÉMICAS DE LA EDUCACIÓN TERCIARIA EN COLOMBIA

### RESUMEN

#### *Clasificación JEL*

I25

J24

D84

#### *Palabras clave:*

Expectativas Subjetivas

Educación en Colombia

Educación terciaria

Analizamos las expectativas académicas y de ingresos de una muestra representativa de estudiantes colombianos en el último año de educación media secundaria, en el cuál consideramos tres posibles escenarios: terminar sus estudios, inscribirse en una carrera vocacional, u obtener un título profesional. Encontramos que los estudiantes tienen expectativas de ingresos correlacionadas con los salarios observados en el mercado laboral. No obstante, sus expectativas de ingreso en el caso de obtener un título profesional son notablemente superiores a los ingresos observados, lo que no ocurre en los otros dos posibles escenarios. También encontramos que las expectativas de ingreso y académicas (resultados de pruebas estandarizadas y la probabilidad de ir a una universidad) están relacionadas con las características socioeconómicas de sus familias, colegio y variables clave del municipio, reflejando aspectos como las brechas de género o el bono de estudiar en un colegio privado. Finalmente, los dos tipos de expectativas están relacionadas con los resultados obtenidos en las pruebas oficiales estandarizadas.

---

\* <https://doi.org/10.32468/espe.8601>

Los autores pertenecen, en su orden, al Departamento de Economía de la Universidad Jorge Tadeo Lozano y a la Facultad de Economía de la Universidad del Rosario.

Correos electrónicos: [luisfw.gamboan@utadeo.edu.co](mailto:luisfw.gamboan@utadeo.edu.co) [paul.rodriguez@urosario.edu.co](mailto:paul.rodriguez@urosario.edu.co)

## 1. Introduction

The purpose of this paper is to assess whether Colombian students in their last year of secondary school have academic and earnings expectations that relate to the labour market, and to the scores of standardised tests used for admission to colleges around the country. We use the case of Colombia, where higher education includes vocational (technical) and academic studies as alternative options. The understanding and practical application of basic principles of science and mathematics is carried out in the former, and its objective is the preparation of graduates for occupations that are classified above skilled crafts, but below scientific or engineering professions. Professional education places a major emphasis on the theories, understanding, and design of new ways to solve problems.

People invest in education under uncertain schemes seeking non-pecuniary goals (Oreopoulos and Salvanes, 2009), but also improving their quality of life by maximising their earning potential. These concepts are summarised in the well-known human capital formation model (Becker, 1964; Mincer, 1974), where earnings expectations play a central role in investment choices.

An important element to consider is the distinction between *ex ante* and *ex post* income variability. In some middle-income and developing countries, that individuals' *ex ante* forecasts about their future income might be below the *ex post* income that they would obtain if they were to invest in education: Jensen (2010) found that teenagers in rural parts of the Dominican Republic do not expect important benefits from secondary education. In a different direction, Hastings et al. (2016) found that students who overestimate or underestimate the costs of college in Chile tend to make “mistakes” regarding human capital formation.

Jensen's results differ from the findings from developed countries. While Dominitz and Manski (1996) found that students know about the existence of a college earnings premium, they demonstrate substantial heterogeneity and commonly overestimate (*ex post*) returns. However, in the case of Duke University students, they are aware of the relative pecuniary benefits of their majors (Arcidiacono et al., 2010). There is also evidence about the role played by externally implemented tests on expected returns, choices, and student motivations (Stinebrickner and Stinebrickner, 2012; Wiswall and Zafar, 2015; Zafar, 2011; Zafar, 2013; Reuben et al., 2017; Sequeira et al., 2016). Based on this evidence, several information-based policy experiments have been conducted around the world. Information on labour market wages around the world has been used in China (Loyalka et al., 2013), Colombia (Bonilla et al., 2015), Finland (Pekkala et al., 2015), India (Sequeira et al., 2016), Mexico (Avitabile et al., 2015) and the United States of America (Fryer, 2013). Others have focused instead on knowledge about costs and financial constraints; for

example, Hoxby and Turner (2013) in the United States of America, and McGuigan et al. (2015) in England. Another methodology, based on reference points, has been used in Madagascar (Nguyen, 2013). The results from these interventions are mixed, with some students updating their beliefs, but often without significant changes in their actual choices. However, there is evidence that another type of intervention, such as conditional cash transfers, might improve students' aspirations (García et al., 2016).

It is important to mention that future earnings expectations may differ from contemporaneous data because of several reasons. This is due to either having private information to explain why their idiosyncratic beliefs differ from the observed mean, or to having incorrect information about such a mean. Systematic differences arise if students consider that economic conditions will be different in the future if there is a generalised error in the information about general wages (as in Jensen's paper), if the learning process starts with a prior belief above or below the mean (optimistic or pessimistic students), or simply if the elicitation instrument induces students to under or over-report their beliefs.

Along with this document, we analyse a nationally representative sample of the last year of upper secondary school students who took the *Saber 11* test in 2013 (the Colombian equivalent to SAT).<sup>1</sup> This is the standardised, mandatory secondary school leaving test in Colombia. Its importance is crucial to students in the future. On one side, it is required to enrol in tertiary education. On the other, it is used as a quality indicator for admission or financial aid in private and public post-secondary institutions. In 2013, several weeks before the test, ICFES (the National Institute for Educational Evaluation) included an additional set of questions in its application form, among which was a subset that included some items about expectations. These questions were to be answered by 10% of the students from the registry (N=58.224 students). They were asked to indicate: (1) their earnings expectations in 5 years, conditional on following a vocational training path,<sup>2</sup> finishing a college degree,<sup>3</sup> or not pursuing any further education above secondary school; (2) their expected results in the mathematics and language tests; and (3) the odds of enrolling in an undergraduate program.

This paper considers (1) how income and academic expectations are related to student, family, school, and

1 Secondary school in Colombia covers grades 6 to 11 in most schools, with grades 9 to 11 being the upper secondary school. Those who graduate from 11 degree obtain a *Bachiller* title, which is equivalent to a high school diploma. Saber Tests (3,5,9, 11 and Pro) are designed and managed by the ICFES (*Instituto Colombiano para la Evaluación de la Educación*).

2 There are two vocational training degrees in Colombia: a technical diploma (1 to 1.5 years), or a technological diploma (2 years). In terms of clarity, we will use the term 'technical' for vocational studies and the term 'professional' for academic programmes studied in universities.

3 Most college degrees in Colombia are between 4 and 5 years long. Hence, we asked for entry wages in the labour market.

municipality characteristics, particularly the observed wages in the local labour market for each student. This is compared with (2) how these expectations are related to test scores and their contribution to our ability to understand the variance in such results. As the questions on expectations were posed in categories, we use an interval regression to read the correlations obtained.

We found that the earnings expectations of Colombian students closely followed the variations in the local labour markets when considering vocational and secondary education scenarios. However, there is a lower correlation between college expected earnings and observed wages in departments. Similar to Bonilla et al.'s (2015) findings, for a sample of schools in Bogotá, students systematically tend to expect college earnings well above those observed for individuals of the same gender, who are 5 years older, and from the same city. Earnings expectations also exhibit characteristics such as gender pay gaps; however, such gaps are not present in earnings for college education. This is linked to the fact that females are more likely to be enrolled in college. Avitabile et al. (2015) demonstrate that females are more likely to increase their schooling effort after being exposed to information related to education and future career prospects.

The expected test scores also present a common gender gap: males expect better results in mathematics than females, while females expect better results in language. Also, students from private schools expect higher test scores, even after considering previous school results in the test. These gaps are several times larger than what can be observed in the data from test scores, which show that preconceptions might play an important role on how individuals make their decisions on human capital formation.

Finally, both earnings and academic expectations are related to observed test scores. However, while the expectations data add private information to the decision process, they do not notoriously improve our understanding of the variance in test results. This suggests, in line with the literature described above on information treatment experiments, that the bottleneck in college formation (human capital accumulation) in middle-income countries is not principally due to an information problem.

The structure of this document is as follows. After this introduction, section two describes the databases and the empirical strategy adopted throughout the paper. This is followed by the results and robustness checks, and, lastly by the conclusion.

## 2. Data and Empirical Exercises

The *Saber 11* Test is designed to be taken by all students in the last year of secondary studies, and it is administered by the Colombian Institute for Evaluation of Education, ICFES. Most of the private universities'

admission procedures are based on the scores from this test or part of it (i.e. Maths scores). During the inscription process, which is carried out about 3 months before the test, students fill out online forms about their personal characteristics and household conditions. For 2013, the online application form included a set of additional compulsory questions for a random sample of 10% of all students (58.224 students) on the two dates of the test (first and second semester).<sup>4</sup> Tables 1 and 2 compare the population and sample means of the principal variables available, and show that the sampled group is balanced in terms of most of the relevant covariates in this study, which will be described below.

To guarantee comparability and avoid *noise* from unobservable factors, the sample was depurated by omitting individuals with missing information or extreme values, for example missing information on parents' education or scores equal to 0 or above 100. More details about these restrictions are presented in the Appendix.

The average student was 16 years old. This reflects that the choice of future investment in education is made by very young people (Tables 1 and 2). At the individual level, there is considerable dispersion in the scores obtained in mathematics and verbal analysis. These test scores were published after the questions regarding expected returns were completed (about 6 months); therefore, expectations are not affected by them. A family's socio-economic index, which is constructed using a set of variables that describe the physical assets owned by the household, reflects that the students' living conditions are equivalent to those of the middle-low class.<sup>5</sup> At the school level, there is a higher incidence of cohorts above 90 students where boys are less numerous than girls. More than 60% of the student population come from households with a low educational background (Table 2). That is, the fraction of students whose mother has a professional education is less than 7.0%, and less than 8.0% in the case of the father. These statistics reflect a similar situation for people around the country. Human capital in parents is crucial in terms of the support provided to enrol in higher education, but also because of its impact on students' motivation. It is interesting to note that the incidence of gold mines is nearly 14%. There are less boys than girls (boys -45.3%), and about 77% come from publicly-managed schools. Another aspect worth mentioning is the low incidence of students with any type of handicap (e.g., visual, auditory, and physical movement), but it is crucial to include it in terms of perceived limitations and expectations for the future.

4 This test is carried out twice per year according to the coexistence of two principal school calendars in Colombia. Those students belong to different educational schedules (88,8 % from "Calendar A" and 11,2 % from "Calendar B").

5 This variable was constructed using the weights defined by the (ICFES, 2010) but excluding parents' education, as its variation will be analysed separately.

Table 1  
Descriptive Statistics I

Continuous Variables	Population				Survey Sample			
	Mean	Mean $\pm$	SD	Min	P10	P50	P90	Max
<b>Student and family level (n=49361)</b>								
Saber 11: Verbal	47.93	47.93	7.67	19.00	38.00	48.00	58.00	95.00
Saber 11: Maths	45.78	45.73	10.63	3.00	34.00	44.00	60.00	100.00
Age	16.68	16.68	1.29	12.00	16.00	16.00	18.00	30.00
Family socio-economic index	18.66	18.60	10.19	0.50	6.88	17.04	32.89	75.78
<b>School level (n=1049)</b>								
Saber 11: Std Mean Maths Scores 2012	45.56	45.45	5.82	28.50	39.63	44.48	52.42	74.42
Saber 11: Std Mean Verbal Scores 2012	46.17	46.14	3.88	35.96	41.53	45.70	51.27	62.10
Saber 11: N. Exam-takers	48.77	46.59*	40.55	2.00	10.00	33.00	102.00	301.00
<b>Municipality level (n=118)</b>								
Wage premium 16 to 11 years of educ.	1.53	1.56	0.30	1.25	1.26	1.45	2.02	2.23
Avg. wage 11 years of educ.	63.40	62.40	11.00	43.84	45.77	63.09	76.56	76.56
MD. Poverty rate	0.69	0.71	0.15	0.30	0.50	0.73	0.90	0.98
Urban-Rural Ratio	3.06	1.84	5.86	0.04	0.16	0.53	3.20	53.56
Homicide rate 2009	27.09	20.20**	25.26	0.00	0.00	11.00	62.00	124.00
Total Population	0.39	0.23	0.51	0.01	0.04	0.11	0.41	3.80
<b>Total Ind.</b>	<b>1.108</b>	<b>118</b>						

† Data from ICFES official classification for 2010 if 2011 is not available.

± Starts show the significance of a test checking if the difference between the sampled and non-sampled mean is equal to zero (\* 0.1 \*\* 0.05 \*\*\* 0.01).

Source: Own calculations based on a 10% student sample from Saber 11 2013-II.

Table 2  
Descriptive Statistics II

Dummy Variables	Population		Survey Sample	
	%	%+-	Ones	SD
<b>Student and family level (n=49361)</b>				
Male	45.66%	45.50%	22.457	49.80pp.
Any handicap	0.30%	0.30%	150	5.50pp.
Father Ed: Less than High School	68.49%	68.21%	33.662	46.57pp.
Father Ed: High School	23.15%	23.58%**	11.640	42.45pp.
Father Ed: Vocational Education	1.53%	1.54%	758	12.30pp.
Father Ed: Professional Education	6.84%	6.67%	3.294	24.96pp.
Mother Ed: Less than High School	62.73%	62.82%	31.002	48.33pp.
Mother Ed: High School	26.65%	26.83%	13.242	44.31pp.
Mother Ed: Vocational Education	1.95%	1.91%	944	13.70pp.
Mother Ed: Professional Education	8.67%	8.44%*	4.166	27.80pp.
<b>School level (n=1049)</b>				
First Semester Calendar	4.24%	4.39%	46	20.49pp.
Private Administration	32.52%	32.32%	339	46.79pp.
Non Mixed Gender	5.07%	4.45%	42	20.64pp.
Saber 11 School Ranking: High and above	42.17%	42.12%	393	49.40pp.
<b>Municipality level (n=118)</b>				
Gold mines	16.88%	20.34%	24	40.42pp.

† Data from ICFES official classification for 2010 if 2011 is not available.

± Starts show the significance of a test checking if the difference between the sampled and non-sampled mean is equal to zero (\* 0.1 \*\* 0.05 \*\*\* 0.01).

Source: Own calculations based on a 10% student sample from Saber 11 2013-II.

We added some additional controls regarding economic and political heterogeneity (Table 3). Among these, some variables relate to economic conditions (poverty rate per municipality and an index of standard of living), the existence of booms (the number of licences for extracting gold, a commodity whose price has boomed since the great recession), and variables related to violence. The existence of mining booms may influence educational expectations, as does the incidence of violent events.

To compare students' expectations with actual earnings, we constructed a distribution of observed earnings in 2013 from the National Household Survey (*Gran Encuesta Integrada de Hogares*, GEIH for its acronym in Spanish). This is based on students' gender, and their age plus 5 years. That is, we have constructed a proxy of their earnings in 5 years. We assume that the observed earnings for 2013 will not change at all within the age, gender, and schooling levels. This is not a particularly strong assumption, given that the macroeconomic conditions in the country have been stable over the last 20 years. The annual inflation rate in Colombia has been below 10% since 2000, and the GDP growth rate has been between 1.7% and 6.9% during the same period.

## 2.1 Subjective Earning Expectations

Our indicator of expected earnings is constructed from the use of the following question in the application form: 'In five years, what monthly income do you think you would earn in each of the following situations (a. You only finish middle education and pursue no more studies; b. You study for a vocational career; c. You finish higher education (any major) at college or university)?' The most useful scenario is one in which perceived labour incomes were gathered as a continuous variable, but ICFES decided to collect it in minimum wage (MW) intervals in order to guarantee time-comparability for questions in future exercises.<sup>6</sup> The minimum wage is an important reference number recognised by people in the Colombian labour market, given that employees typically earn either 1 MW, or their wage is defined in terms of MW. It is also used for defining the value of several official figures, such as fines and legal compensations. The categories selected by the ICFES were:

1. *Less than 1 MW,*
2. *Between 1 and 2 MW,*
3. *Between 3 and 4 MW,*

Table 3  
Data Sources

Variables	Year	Source	Description
<b>Municipality</b>			
Quality of Life	2005	DNP	<i>Índice de Condiciones de Vida</i> . PCA based index which summarizes human and physical capital characteristics
MD. Poverty rate	2005	DNP	Percentage of people in poverty according to the Multidimensional Poverty Index
Homicide rate*	2009	Medicina Legal†	Per 100.000 inhabitants
Total Population*	2005	DANE	Total residing population over 100.000 from the National Census
Urban-Rural Ratio*	2005	DANE	Based on the National Census
Total Attacks	2005-2009	CINEP	Total number of attacks
Gold mines	2000-2007	SGC	Total mining rights
Coal mines	2000-2007	SGC	Total mining rights
<b>Department</b>			
Avg. wage 11 years of educ.	2013	GEIH	Average monthly wage conditional on having 11 years of education
Wage premium 16 to 11 years of educ.	2013	GEIH	Difference of wages conditional on having 16 and 11 years of education, over 11 years of education wages

\*Obtained via SIGOT web page. † National Legal Medicine and Forensic Sciences Institute.  
 DANE: National Statistics Department. DNP: National Planning Department.  
 SGC (INGEOMINAS): Colombian Geological Service. CINEP: Popular Education and Research Centre.  
 Source: Authors' calculations

6 For 2013, the minimum wage earned in Colombia was COP 589,500 equivalent to USD 320. But it is important to mention that the employer has to pay an additional amount for social security, which increases the salary to USD 470 in the case of formal workers.

4. *Between 5 and 7 MW,*
5. *Between 8 and 10 MW,*
6. *More than 10 MW.*

The choice of intervals always includes a subjective judgment because of the *trade-off* between the number of intervals and their amplitude. Fewer intervals make it harder to detect variations on the conditional mean of the latent variable, but make it simpler for the student to understand and answer the question. Moreover, the values defined for some categories were misleading, as they avoid the possibility of some intervals (e.g. between 2 and 3 MW), and are completely arbitrary. Although these are the most obvious problems (which we address in the robustness checks by modifying the values of the cut-offs), there are conceptual issues that are harder to tackle. In particular, the fact that we are imposing values for the thresholds might generate *anchoring*: that is, individual responses might be determined partly due to the way in which the options are presented. In the case of education, Heckman et al. (2010) suggest it is desirable to anchor skills in a common scale in order to gain a better understanding of the relative importance of cognitive skills and their importance of investments at different stages of the life cycle; however, the number of categories and their meaning provide information to the respondent that might influence their choice. In other words, if someone is questioned about their expected performance in a continuous (i.e., categorical) scale, his/her answers may differ because the structure of the options indicates a relative perceived self-valuation.<sup>7</sup> Hence, even though we present some numbers that can be interpreted in terms of the Colombian currency, they are mainly for illustrative purposes. In fact, our analysis focuses on terms of order, rather than magnitude.

## 2.2 Test Scores and Progress to Higher Education Expectations

Among other factors, test performance depends on *self-motivation* and expectations about future educational choices. One of the advantages of our sample is the availability of information about the scores expected by the students, as well as whether they expect to be enrolled in a tertiary education programme. These expectations provide us with some intuition about self-perceived ability, tastes and effort.

Regarding mathematics and language scores, students were asked to choose their expected results from the categories below. For our analysis, the cut-off points are translated into standardised test scores.<sup>8</sup>

7 See Attanasio et al. (2005); Attanasio (2009); Delavande et al. (2011) for a deeper discussion of the main particularities of eliciting subjective survival probabilities.

8 The corresponding cut-offs are around -6.1, -2.2, 0.4, 3, 6.8 standard deviations for both the mathematics and language tests.

1. *Very Low: 0-30 points,*
2. *Low: 30.01-50 points,*
3. *Medium: 50.01-70 points and*
4. *High: 70 points and above*

The Likert Scale was employed in expectation of enrolment in tertiary education.

- a. *I will not continue studying*
- b. *It is unlikely that I will enter a higher education program*
- c. *It is likely that I will enter a Higher education program and*
- d. *Surely I will enter a Higher Education program*

For our empirical exercises, these answers were collapsed in a dummy variable that it is equal to 1 for *c* and *d*, and 0 otherwise.

## 2.3 Empirical Strategy

We proceed to carry out two types of exercises according to the database. First, we check the existence of any correlations between expectations and some of the socioeconomic variables. Second, we intend to study the relationship among test scores and expectations. In the first case, the structure of the database only allows us to observe labour income and test scores as categorical variables. Then, we proceed to use interval regression to determine how the observed characteristics of the students, their families, and their schools and municipalities are related to expectations.

Apart from reported wages  $w_{ijl}$  for student  $i$  from school  $j$  in level  $l$ , earnings ratios between wage with schooling level  $l$  over wage in  $m$  level ( $R_{i,l,m}$ ) are calculated by assuming the midpoint of each reported wage category.

Under the Interval Regression approach, the underlying variable  $w^*$  is observed in  $C$  intervals defined by the cut-offs  $\{\kappa_1, \dots, \kappa_{C-1}\}$ . Then, if  $w^*$  is between the values  $\kappa_{c-1}$  and  $\kappa_c$ , the categorical counterpart will take the value  $w^c$  (see equation 1).

$$\begin{aligned} w_{ijl}^* &= \theta Z_{ijl} + u_{ijl} \\ &= \sum_{l=0}^2 (r_l + r_{lx} X_{ij} + r_{lm} M_j S_{ijl} + e_{ijl}) \end{aligned} \quad (1)$$

By following Stewart's (1983) approach, it is necessary to impose a restriction on the way the truncation is done. That is, we need to model the probability of an individual choosing a particular category  $w^c$ , as in equation 2.

$$\begin{aligned} \Pr(w_{ijl}^c) &= \Pr(\kappa_{c-1} \leq w_{ijl}^* \leq \kappa_c) \\ &= \Pr(\kappa_{c-1} - \theta \cdot Z_{ijl} \leq e_{ijl} \leq \kappa_c - \theta \cdot Z_{ijl}) \\ &= F_e(\kappa_c - \theta \cdot Z_{ijl}) - F_u(\kappa_{c-1} - \theta \cdot Z_{ijl}) \end{aligned} \quad (2)$$

For simplicity, we can assume  $F_e \sim N(0, \sigma^2)$ , so the model can be easily estimated by maximum likelihood. As we assume that the category thresholds are credible, the

marginal effects can be understood in the usual way. The assumption of normality is crucial for providing monetary-based estimates on the marginal effects of each regressor. Hence, such values might change as another distribution is considered. However, the relationship between regressors and beliefs is given by the ordinal feature of our perceived returns variable.

Consistency is achieved if we assume that  $e_{ijt}$  is uncorrelated with respect to  $X_{ijt}$ . This might be a strong assumption, given that unobserved characteristics such as motivation might be related to better academic achievement and to higher expected returns. This would be reflected in an upward bias of our estimates. Additionally, the values of our estimates in terms of ‘pesos’ are determined to a large extent by the values of the thresholds. How would the results change if the cut-offs were chosen differently? This is provided by the normality assumption of  $e_{ijt}$ .

In the second case, test scores and expectations, our dataset allowed us to compare both earnings and academic expectations with the actual test scores, which took place several months after the elicitation of beliefs. We can thus understand how private information embedded in expectations is related to test performance. To assess their relationship, we will estimate a linear regression for test scores, and discuss whether expectations information adds new information into the analysis of observed characteristics. Although school-level fixed effects are normally considered in this type of exercises because of the effect of unobserved variables on individual beliefs, the main issue is related to maximum likelihood estimation. The number of schools (9888 in the dataset) makes it hard for the optimisation algorithm to estimate such large number of parameters.<sup>9</sup> This is an important caveat for the individual and household level regressors, where the school and municipality levels might not be enough for disentangling individual from classroom effects.

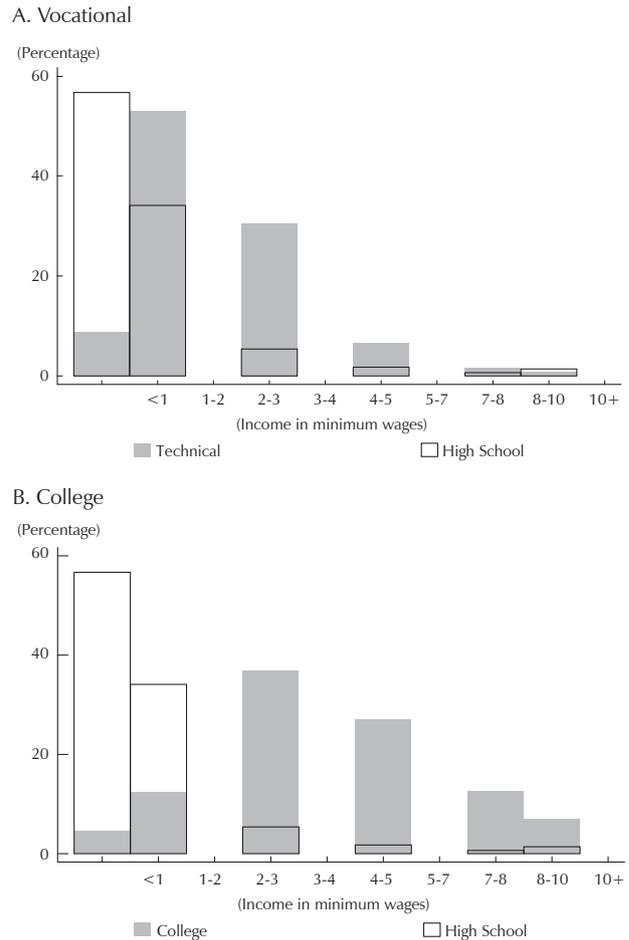
### 3. Results

#### 3.1 Earnings Expectations and Observed Labour Income

Our analysis allows us to compare earnings expectations between two scenarios, high school vs. vocational studies, or high school vs. college. A first result is that students consider that investing in additional years of education will provide a substantial payoff in the case of college, and notoriously less in the case of vocational

education. This is observed in Figure 1, where both the distribution of expected earnings from vocational and college education are to the right of expected earnings from high school distribution.

Figure 1  
 Distribution of Subjective Expected Earnings

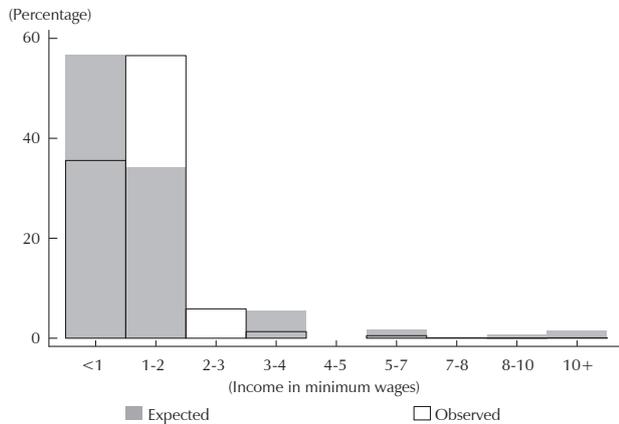


Source: Own calculations based on ICFES expectations survey.

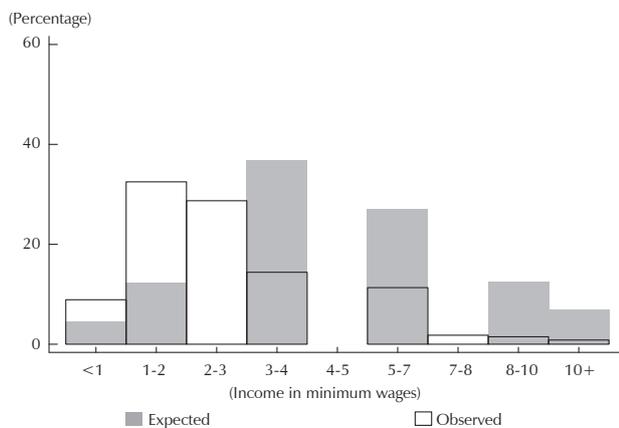
Perhaps the most striking result comes from comparing the current labour market information and students’ expectations. Colombian students do not underestimate college graduates’ wages. Instead, they highly overestimate them. Figure 2 compares the expected and observed earnings for 2013, which are conditional on attaining a college degree or just completing high school. There are several points to consider in these figures. First, it is likely that part of the overestimation comes from the anchoring: students tend to value college education in a higher income category just because they perceive them as “high”. Second, since the question design omitted some categories, the central values of our distribution are biased. Third, overconfidence is a common result when eliciting returns on investments, as discussed by Brunnermeier and Parker (2005). As a result, it is hard to say by how much the students are overestimating their potential future earnings.

<sup>9</sup> In these cases, is a challenging computational problem to fit the model. In this case, there is no suitable transformation to avoid estimating them. On top of this practicality, there is evidence of problems with the statistical properties of the estimators when an important number of fixed effects on non-linear models, even with standard probits (Greene, 2002).

Figure 2  
 Expected vs. Observed Earnings  
 A. High School Only



B. College

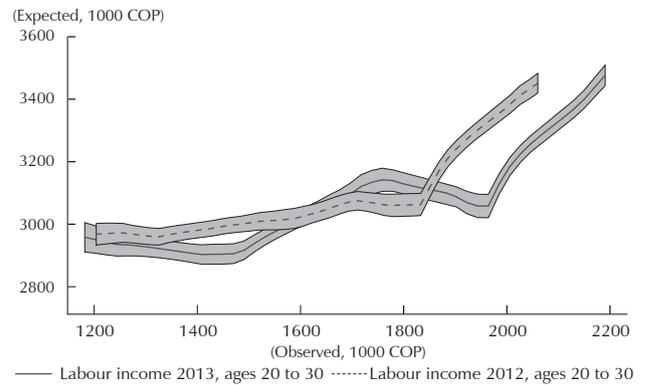


Note: Observed wages correspond to monthly earnings reported by individuals aged 20 to 30 years old who are working and meet the education level criteria in the GEIH survey. Such figures were divided by 589.500 COP, an official monthly minimum wage in 2013, and the classified into the corresponding categories.  
 Source: Own calculations based on the GEIH and ICFES expectations survey.

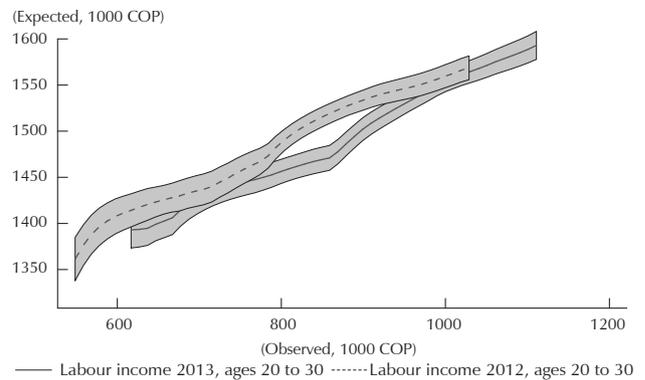
Despite these limitations, as shown in the local linear analysis in Figure 3, expectations follow variations in observed earnings (both 2012 and 2013 values) of people aged 20 to 30 years old.<sup>10</sup> However, the relationship is flat for college in places where average earnings are between COP 1.2 and 1.6 million (around 2 to 2.5 minimum wages). There are several potential explanations for this. First, it could be a result of the categorical design of the question, which makes it harder to detect small variations. Second, it could be due to a lack of diffusion of wage information in areas where wages are low. Third, it could indicate that students consider that the relevant labour market is not the same as the one in which they live. This is plausible, considering that many students typically migrate to the main cities to pursue college education.

10 For this exercise, each category of expected earnings was assumed to be represented by its mean point. While this assumption is crucial for the values on the Y axis, it is irrelevant for the direction of the correlation.

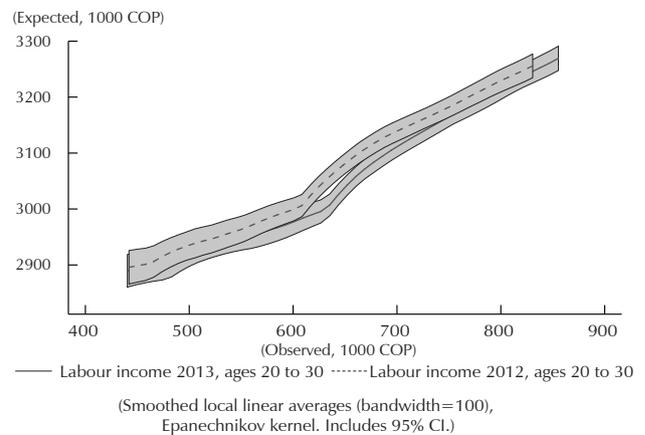
Figure 3  
 Observed and Expected Earnings by level of Education  
 A. College



B. Vocational



C. High School



Source: Authors' own calculations.

Table 4 reports the result of the interval regressions analysis. The first three columns summarise the marginal coefficients for each academic level expected, while the last two include earnings ratios (the ratio of perceived earnings between one academic level and the prior level). From this table, we can comment on some interesting correlations.

Table 4  
 Interval Regression Marginal Effects

	A. Income Levels			B. Ratio Rs,0	
	High School	Vocational	Professional	Vocational	Professional
<b>Personal and Family Characteristics</b>					
Age	6.03 (3.72)	7.81** (3.85)	-9.10 (6.34)	-0.50 (0.85)	-4.28** (1.97)
Male	37.75*** (9.10)	56.29*** (9.59)	14.96 (15.56)	-2.17 (2.05)	-17.46*** (4.83)
Any handicap	-160.44*** (46.06)	-195.41*** (69.28)	-548.93*** (125.16)	14.91 (11.63)	11.14 (29.30)
Family socio-economic index	7.50*** (0.64)	11.54*** (0.66)	25.18*** (1.06)	-0.38*** (0.14)	-0.58* (0.34)
Father Ed: Less than High School	-8.16 (11.65)	-4.33 (12.02)	13.67 (19.56)	1.49 (2.64)	5.97 (6.19)
Father Ed: Vocational Education	92.41** (44.37)	45.89 (41.68)	22.79 (64.31)	-17.25* (9.78)	-44.54* (23.49)
Father Ed: Professional Education	-18.51 (20.69)	26.23 (21.66)	106.11*** (34.50)	8.06* (4.67)	23.30** (10.99)
Mother Ed: Less than High School	-3.73 (11.19)	4.56 (11.62)	20.12 (18.87)	1.53 (2.53)	4.53 (5.94)
Mother Ed: Vocational Education	22.06 (36.81)	-24.31 (37.22)	17.89 (59.67)	-8.70 (8.11)	-9.01 (19.61)
Mother Ed: Professional Education	-40.05** (19.18)	-80.22*** (19.10)	26.32 (31.16)	-0.36 (4.34)	24.00** (10.24)
<b>School characteristics</b>					
First Semester Calendar	227.87*** (45.60)	366.39*** (56.48)	250.82*** (79.08)	-9.63 (9.56)	-84.53*** (23.50)
Private Administration	39.99*** (12.89)	23.39* (13.58)	37.41* (22.08)	-7.01** (2.81)	-15.69** (6.81)
Saber 11 School Ranking: High and above	-39.26*** (12.77)	-30.35** (13.27)	4.77 (21.85)	5.92** (2.89)	20.80*** (6.80)
Saber 11: Std Mean Maths Scores 2012	0.51 (1.99)	7.74*** (2.07)	22.33*** (3.32)	0.88** (0.44)	2.64** (1.06)
<b>Municipality characteristics</b>					
Wage premium 16 to 11 years of educ.	14.35 (26.29)	-6.10 (26.78)	4.02 (44.65)	-4.40 (6.03)	-6.85 (14.02)
Avg. wage 11 years of educ.	1.25* (0.74)	1.57** (0.74)	-1.32 (1.21)	-0.11 (0.16)	-0.82** (0.39)
MD. Poverty rate	52.71 (34.34)	2.22 (34.55)	-13.34 (56.61)	-12.95* (7.74)	-28.83 (18.23)
Urban-Rural Ratio	0.16** (0.08)	0.07 (0.08)	-0.25** (0.12)	-0.03* (0.02)	-0.11*** (0.04)
Homicide rate 2009	0.03 (0.14)	-0.30** (0.14)	-0.53** (0.24)	-0.05 (0.03)	-0.09 (0.08)
Total Population	-1.36** (0.62)	0.82 (0.62)	5.28*** (1.00)	0.45*** (0.14)	1.39*** (0.33)

Table 4 (continued)  
Interval Regression Marginal Effects

	A. Income Levels			B. Ratio Rs,0	
	High School	Vocational	Professional	Vocational	Professional
Gold mines	-30.74** (13.22)	-29.95** (13.82)	-75.47*** (22.92)	3.83 (2.91)	5.99 (6.92)
Constants		174.89***	123.88***		
Number of Obs.	133287				
Number of Indiv.	44429				
LR $\chi^2$ (74) test [p-val]	77005.6967 [ 0.00]				
Log-likelihood	-266652.9420197413				
AIC	533457.88				
BIB	534202.70				
Percentage of LB hits / UB hits	65.12 / 64.29				

Calculated by the Delta method at the averages. SE clustered at individual level. Significance: 99% \*\*\*, 95% \*\*, 90% \*.  
Source: Authors' own calculations.

### 3.1.1 Individual and family characteristics

First, older students expect lower earnings from investing in professional studies. This can be explained by the fact that older students are both more likely to have repeated a year, or to have suffered from gaps during their secondary education. These two elements are related to both socio-economic conditions and cognitive skills, which are directly related to the earnings ratio.

A second relevant result is related to gender. Differences in perceived earnings are common for lower educational investments. Boys tend to expect higher wages from basic and vocational education than girls, but this difference disappears for college-level income. This is an interesting point for future research because it could be associated with labour market barrier beliefs in women regarding the availability of vocational jobs. Previous findings suggest that parents' expectations matter in terms of investing in education, and they are strong predictors for the probability of being enrolled in secondary school (Attanasio and Kaufman, 2009; Armand, 2015). Recognising that 'drop out' levels were still important, the Central government implemented conditional cash transfers such as *Familias en Acción* and *Jovenes en Acción*. These benefits were intended to reduce demand barriers to human capital accumulation, which increased the number of girls completing high school and enrolling in higher education. These initiatives increased girls' competences and relative labour income. This contributed to a decline in perceived and observed gender wage gaps during the last years in Latin-American countries (Ñopo and Hoyos, 2010; Barrera-Orsorio et al., 2008). Another common earnings gap is noted among people who report a disability, who notoriously expect lower wages in any category.

Third, it seems that more privileged students tend to value higher earnings at each additional educational level. That is, students belonging to wealthier families, or with college-educated parents, will opt to increase their human capital more than those who do not have enough economic resources.

A significant result is that students whose parents have a vocational education report higher earnings for secondary-only education, which means that the perceived benefits of both vocational and college education are lower than for those children with parents without post-secondary education. There might be at least two reasons explaining this result. On one hand, students from low socioeconomic background could have lower expectations because of the observed realizations in their environment that downplay their aspirations. On the other hand, more informed youths tend to be less optimistic about the added value of additional investments on education.

### 3.1.2 School characteristics

School characteristics are essential for explaining both income and academic expectations, not only because schools train children for the labour market and tertiary education, but also because they reflect parents' choices. Children enrolled in good schools evidence the importance (and financial means) that parents give to education, which is probably correlated with the pecuniary returns on such an investment.

The first element to consider is the quality of the school. We measure it according to prior results on the *Saber 11*, and on the ranking given by schools. They show that better schools imply higher earnings if the student goes to college, and lower ones for any other level.

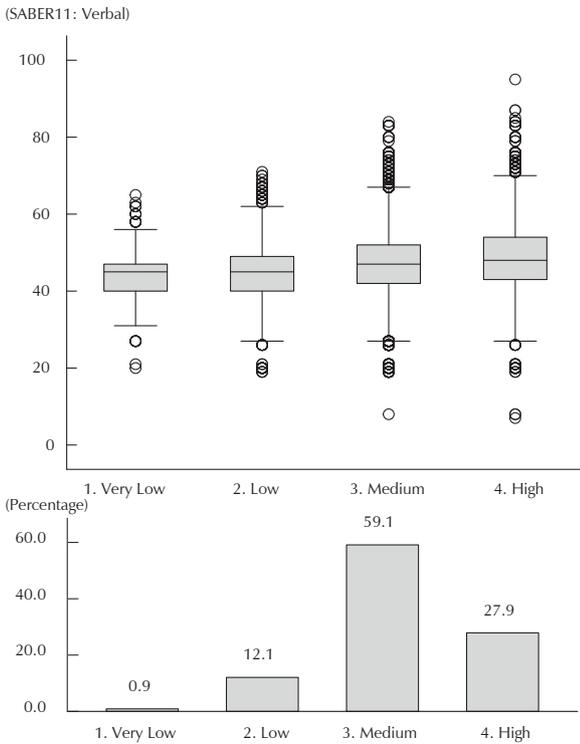
A second aspect to consider is whether a student is enrolled in a public or private school. Public schools are subsidised, but, more importantly, they have fewer tools for controlling the background of their students at enrolment. Private schools can target specific needs of families (such as religion), filter their students with entry exams, and choose the socio-economic background of their students by determining their tuition fees. Results show that students from private schools believe that they will be *better-off* in any scenario by approximately the same amount of money, which slightly reduces both the college and vocational students' ratio. This result is particularly interesting, as it is already conditional on the socio-economic characteristics of the household, municipality characteristics, and school ranking. Since private schools are highly heterogeneous in terms of the economic conditions faced by the students and they cover about 30% of total enrolment, their expectations may differ because of multiple reasons. Schools that apply the *Saber* test in the first semester are private and represent a small fraction of total enrolment (less than 5%), but they provide education to the upper tail of income distribution and their students expect a profitable future, on average.

### 3.1.3 Municipality characteristics

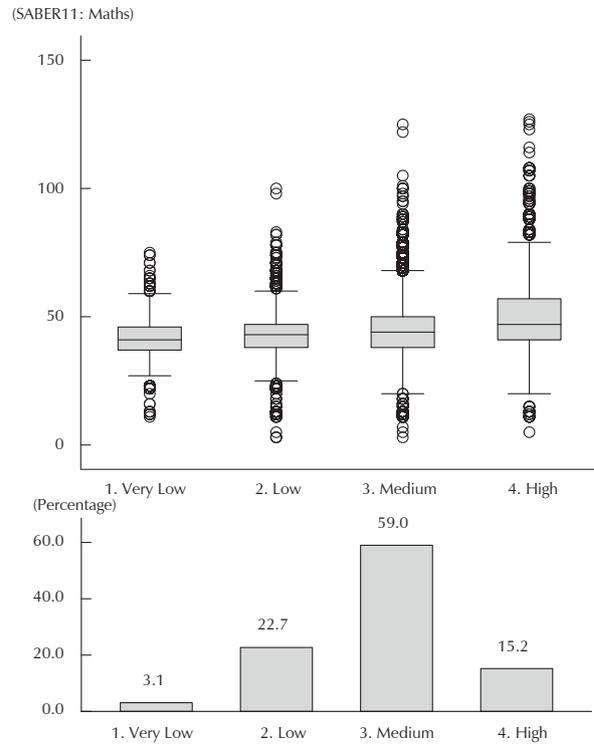
Once we consider personal, family and school variation, the role of municipality characteristics is notoriously less important. Elements such as: poverty rate, urban-rural ratio, and homicide rate are irrelevant once we consider wages and population size. The table shows that students in bigger cities expect higher salaries after college education. It also reflects the lower earnings ratio of college and vocational studies in places that pay better wages to workers with secondary-education. This reflection of regional labour market characteristics, also discussed above, shows that individuals are informed about wages, despite the general overestimation of tertiary education earnings.

A particularly interesting case is the presence of gold mines. Whether mining booms undermine expected returns to education is an open research question that is relevant for developing countries, but which cannot be addressed within this framework. However, the present results suggest that further analysis of this question is relevant: the presence of a gold mine in the municipality is associated to lower income expectations in any educational attainment scenario. Nevertheless, gold-

Figure 4  
 Test Realisations and Expectations  
 A. Language  
 Expect results in Language to be



B. Mathematics  
 Expect results in Mathematics to be



Source: Own calculations based on a sample from SABER 11 2013 data.

mine municipalities have families with lower income, in general, according to the *Saber 11* socio-economic index,<sup>11</sup> which could confirm this finding.

### 3.2 Expected Test Scores

Figure 4 compares expectations about test scores and actual results. On average, expectations clearly signal performance. There is a significant difference in mathematics of around 9 points (0.8 standard deviations) between those who reported the highest category against those who reported the lowest. In the case of the language test, this difference is about 5 points (0.7 SD) and significant at the 99% level.

Additionally, students who expect longer educational investments (who are more likely to progress into higher education) exhibit higher performance. Figure 5 shows that the distribution shifts with each category. For both language and mathematics, there is a significant difference between the lowest and highest categories. These results seem to suggest that there are important links among

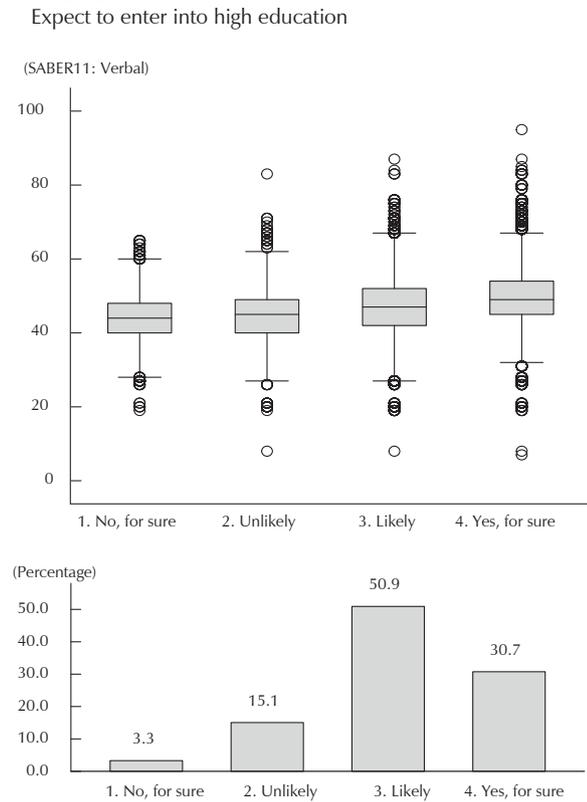
these variables, and that both relationships encapsulate information on beliefs, ability, and other inputs that are crucial for the performance of test-scores.

A more formal analysis is presented in Table 5, which exhibits the marginal effects after an interval regression. The table shows that the key predictors of test results such as age, socio-economic index, parental education, school quality, or the presence of a disability are correlated with both test results and education attainment expectations. There are clear differences between boys and girls in terms of expected test results and the probability of pursuing tertiary education. Also, as with expected earnings, students from private schools expect higher test scores and attendance at college.

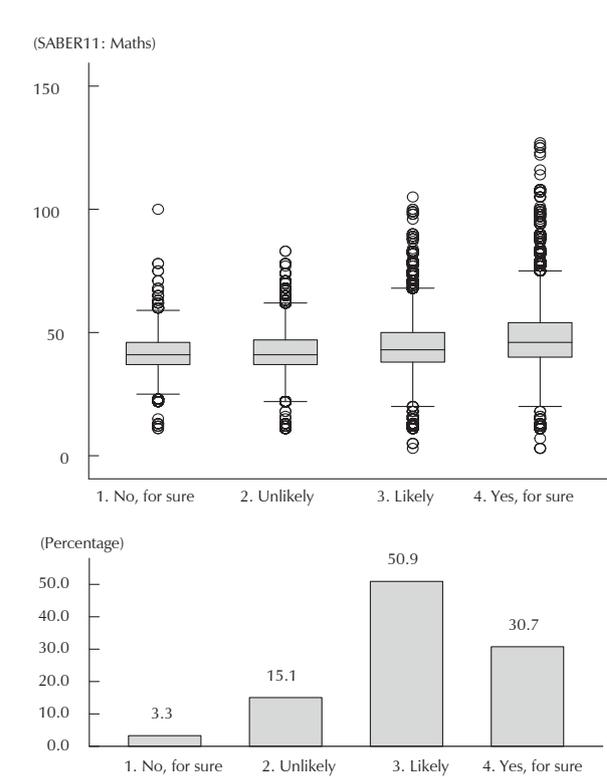
### 3.3 Test Results and Expectations

Table 6 summarises the relationship between test scores and both expected earnings and academic achievement. In this case, the analysis is based on a linear regression where the dependent variables are the standardised mathematics and language results. Columns

Figure 5  
 Test Scores and Expectation to Progress into Tertiary Education  
 A. Language



B. Mathematics



Source: Own calculations based on a sample from SABER 11 2013 data.

11 These are 185 out of 1095 municipalities with available information in the sample, with an average socio-economic income index of 0.23 standard deviations (significant at 99% level) below non-gold-mines municipalities' average.

Table 5  
 Expected academic outcomes and student, school and municipality characteristics

Independent Variables	$\bar{X}$ (1)	Language (2)	Mathematics (3)	College Likely (4)
<b>Personal and Family Characteristics</b>				
Age	16.68 ( 1.29)	−0.0423*** (0.0076)	−0.0459*** (0.0080)	−0.0166*** (0.0014)
Male	0.45 ( 0.50)	−0.3067*** (0.0169)	0.2093*** (0.0183)	−0.0262*** (0.0035)
Any handicap	0.00 ( 0.06)	−0.5609*** (0.1826)	−0.5666*** (0.1613)	−0.0474* (0.0284)
Family socio-economic index	18.60 ( 10.19)	0.0072*** (0.0012)	0.0040*** (0.0013)	0.0072*** (0.0003)
Father Ed: Less than High School	0.68 ( 0.47)	−0.0916*** (0.0213)	−0.0595*** (0.0227)	−0.0242*** (0.0047)
Father Ed: Vocational Education	0.02 ( 0.12)	−0.0771 (0.0655)	−0.0147 (0.0717)	0.0017 (0.0161)
Father Ed: Professional Education	0.07 ( 0.25)	−0.0062 (0.0359)	0.0153 (0.0384)	0.0411*** (0.0107)
Mother Ed: Less than High School	0.63 ( 0.48)	−0.0160 (0.0207)	−0.0495** (0.0219)	−0.0181*** (0.0045)
Mother Ed: Vocational Education	0.02 ( 0.14)	0.0726 (0.0591)	0.1163* (0.0636)	0.0251 (0.0163)
Mother Ed: Professional Education	0.08 ( 0.28)	0.1203*** (0.0338)	0.1399*** (0.0361)	0.0312*** (0.0092)
<b>School characteristics</b>				
First Semester Calendar	0.04 ( 0.19)	−0.2081 (0.1304)	−0.0182 (0.1360)	−0.0103 (0.0265)
Private Administration	0.26 ( 0.44)	0.1899*** (0.0345)	0.2400*** (0.0374)	0.0194*** (0.0075)
Saber 11 School Ranking: High and above	0.50 ( 0.50)	0.0981*** (0.0327)	−0.0045 (0.0353)	−0.0021 (0.0064)
Saber 11: Std Mean Maths Scores 2012	46.51 ( 5.39)	−0.0039 (0.0048)	0.0347*** (0.0052)	0.0032*** (0.0010)
Saber 11: Std Mean Verbal Scores 2012	46.92 ( 3.58)	0.0494*** (0.0085)	0.0199** (0.0092)	0.0067*** (0.0017)
Saber 11: N. Exam-takers	90.36 ( 86.05)	−0.0002 (0.0002)	−0.0002 (0.0002)	−0.0000 (0.0000)
Non Mixed Gender	0.07 ( 0.26)	−0.0015 (0.0494)	−0.0005 (0.0495)	−0.0009 (0.0138)
<b>Municipality characteristics</b>				
Wage premium 16 to 11 years of educ.	1.60 ( 0.25)	0.0134 (0.0759)	0.0382 (0.0792)	0.0848*** (0.0147)
Avg. wage 11 years of educ.	66.59 ( 11.37)	−0.0030 (0.0019)	−0.0062*** (0.0020)	−0.0004 (0.0004)
MD. Poverty rate	0.47 ( 0.21)	−0.0528 (0.0895)	0.2755*** (0.0970)	−0.0534*** (0.0166)

Table 5 (continued)  
Expected academic outcomes and student, school and municipality characteristics

Independent Variables	$\bar{X}$ (1)	Language (2)	Mathematics (3)	College Likely (4)
Urban-Rural Ratio	104.63 (171.62)	0.0009*** (0.0002)	0.0007*** (0.0002)	0.0000 (0.0000)
Homicide rate 2009	38.02 (33.58)	0.0000 (0.0004)	0.0004 (0.0004)	0.0001 (0.0001)
Total Population	16.11 (25.36)	-0.0053*** (0.0016)	-0.0051*** (0.0017)	-0.0011*** (0.0003)
Gold mines	0.12 (0.33)	-0.1585*** (0.0357)	-0.1336*** (0.0365)	-0.0138** (0.0070)
N Observations		44429	44429	44429
N Clusters		8446	8446	8446

Columns 1 and 2 present marginal effects from a model using intervals of the expected *Saber 11* test scores as outcomes. Column 3 presents marginal effects after a Probit regression where the outcome is whether the student considers it likely or very likely to enrol into college education. SE clustered at school level. Significance: 99% \*\*\*; 95% \*\*; 90% \*. Source: Authors' own calculations.

2 and 4 present a regression without expectations, while Columns 3 and 5 include such information.

In terms of the basic individual, family, school, and municipality characteristics, the pattern is the same as for the expectation questions. A difference that is worth mentioning is the gender gap in language. While in mathematics there is a clear difference between boys and girls both in terms of expectations and results, boys expect lower results than girls in the language test. This difference is not observed in the actual data.

Another interesting difference is the role of private schools. Conditional on socio-economic characteristics and prior school results, students in private schools expect better scores than those in public ones. However, in the data analysed data, they do slightly worse than public ones. By dividing private schools according to their tuition fees (lower 50% and upper 50%), we found that the increase in earnings expectations is concentrated on the upper tuition group, and the reduction on language test scores is found for the lower tuition group. Nevertheless, the upper tuition level does not perform better in the test results once all the controls and expectations information are considered.

Another result is the explanatory power of expectations. Excluding expectational information, we were able to explain around 23% of the performance in mathematics and 25% of the language results. Once we included beliefs data, the mathematics  $R^2$  remained at 23% and language increased to 27%. This result tells us that expectations do not improve our understanding of the test results, once we consider the main individual, household, school, and municipality covariates. In other words, information on items such as motivation, self-esteem, and non-cognitive abilities, which are contained in the expectations, does not improve the ability of the model to explain the variation on test scores. This does not mean that such variables are not relevant, rather than test-score expectations do

not capture such information appropriately. There is an important caveat, however: that the categorical responses used for eliciting information might be responsible for the loss of relevant information. Moreover, with our current design, we do not know whether this information improves elements such as actual enrolment in tertiary education.

### 3.4 Robustness Checks

We performed several robustness checks to assess the correlations described about earnings expectations. A more detailed description and tables with their outputs are available as an online appendix.

First, we redefined the cut-offs of the interval regressions to consider whether the marginal effects were affected by the numbers imposed by the upper and lower limits stated for each category. The purpose of this exercise was to remove the gaps between the non-continuous intervals by establishing a cut-off between them: for two given intervals  $[A, B]$  and  $[C, D]$  where  $D > C > B > A$ , we considered defining as the cut-off between categories first  $B$ , then  $C$ , and finally  $\frac{B+C}{2}$ . Marginal effects remained the same in sign and significance for all variables, which is the goal of our estimation strategy given the limitations imposed by the data collection exercise.

Second, we considered the inclusion of individual and household covariates alone, as well as how the coefficients vary when school and municipality level controls were considered. The purpose of this exercise was to detect if the main effects were sensitive to the specification. As a result, the magnitude of some of the individual and household coefficients increased when the school and municipality controls were excluded, but again, sign and significance were preserved. Also, we experimented with alternative

Table 6  
 Subjective expected earnings, academic outputs, and Test Scores

Independent Variables	$\bar{X}$ (1)	Language		Mathematics	
		(2)	(3)	(4)	(5)
Ratio College over HS	6.81 ( 5.37)		0.0037*** (0.0009)		0.0045*** (0.0012)
Ratio College over Tech	2.47 ( 1.92)		0.0044* (0.0024)		0.0011 (0.0032)
Reports to be likely to enrol into tertiary education	0.84 ( 0.37)		0.0821*** (0.0117)		0.0514*** (0.0151)
Expected low result	0.22 ( 0.41)				0.1105*** (0.0342)
Expected medium result	0.60 ( 0.49)				0.2702*** (0.0335)
Expected high result	0.16 ( 0.37)				0.6577*** (0.0373)
Expected low result	0.11 ( 0.31)		0.1038** (0.0476)		
Expected medium result	0.59 ( 0.49)		0.1954*** (0.0464)		
Expected high result	0.29 ( 0.46)		0.3153*** (0.0470)		
<b>Personal and Family Characteristics</b>					
Age	16.68 ( 1.29)	-0.0813*** (0.0040)	-0.0773*** (0.0040)	-0.0933*** (0.0052)	-0.0873*** (0.0051)
Male	0.45 ( 0.50)	0.0316*** (0.0085)	0.0483*** (0.0085)	0.4243*** (0.0117)	0.4080*** (0.0115)
Any handicap	0.00 ( 0.06)	-0.3858*** (0.0806)	-0.3582*** (0.0786)	-0.5755*** (0.1051)	-0.5227*** (0.1065)
Family socio-economic index	18.60 ( 10.19)	0.0069*** (0.0005)	0.0060*** (0.0006)	0.0080*** (0.0008)	0.0072*** (0.0008)
Father Ed: Less than High School	0.68 ( 0.47)	-0.0030 (0.0107)	0.0024 (0.0107)	0.0305** (0.0144)	0.0360** (0.0143)
Father Ed: Vocational Education	0.02 ( 0.12)	0.0054 (0.0354)	0.0083 (0.0351)	0.1260*** (0.0481)	0.1248*** (0.0466)
Father Ed: Professional Education	0.07 ( 0.25)	0.1194*** (0.0189)	0.1174*** (0.0188)	0.1533*** (0.0269)	0.1504*** (0.0263)
Mother Ed: Less than High School	0.63 ( 0.48)	-0.0110 (0.0100)	-0.0084 (0.0100)	0.0136 (0.0141)	0.0180 (0.0139)
Mother Ed: Vocational Education	0.02 ( 0.14)	0.0591* (0.0303)	0.0540* (0.0301)	0.0785* (0.0445)	0.0658 (0.0439)
Mother Ed: Professional Education	0.08 ( 0.28)	0.1294*** (0.0172)	0.1213*** (0.0172)	0.1545*** (0.0249)	0.1376*** (0.0244)

Table 6 (continued)  
Subjective expected earnings, academic outputs, and Test Scores

Independent Variables	$\bar{X}$ (1)	Language		Mathematics	
		(2)	(3)	(4)	(5)
<b>School characteristics</b>					
First Semester Calendar	0.04 (0.19)	0.0879 (0.0569)	0.1040* (0.0575)	0.0130 (0.0652)	0.0203 (0.0649)
Private Administration	0.26 (0.44)	-0.0294** (0.0127)	-0.0364*** (0.0127)	0.0249 (0.0177)	0.0026 (0.0177)
Saber 11: Std Mean Verbal Scores 2012	46.92 (3.58)	0.0567*** (0.0034)	0.0539*** (0.0034)	0.0203*** (0.0050)	0.0179*** (0.0050)
Saber 11: N. Exam-takers	90.36 (86.05)	-0.0000 (0.0001)	-0.0000 (0.0001)	-0.0001* (0.0001)	-0.0001* (0.0001)
Non Mixed Gender	0.07 (0.26)	0.0857*** (0.0200)	0.0875*** (0.0201)	0.1076*** (0.0310)	0.1081*** (0.0306)
<b>Municipality characteristics</b>					
Wage premium 16 to 11 years of educ.	1.60 (0.25)	0.0515** (0.0256)	0.0422* (0.0256)	-0.0601 (0.0370)	-0.0736** (0.0371)
Avg. wage 11 years of educ.	66.59 (11.37)	0.0035*** (0.0007)	0.0037*** (0.0007)	-0.0008 (0.0010)	-0.0001 (0.0010)
MD. Poverty rate	0.47 (0.21)	-0.1114*** (0.0317)	-0.0976*** (0.0317)	0.0439 (0.0450)	0.0368 (0.0452)
Urban-Rural Ratio	104.63 (171.62)	-0.0001 (0.0001)	-0.0001** (0.0001)	0.0000 (0.0001)	-0.0000 (0.0001)
Homicide rate 2009	38.02 (33.58)	0.0001 (0.0001)	0.0001 (0.0001)	-0.0004* (0.0002)	-0.0004** (0.0002)
Total Population	16.11 (25.36)	-0.0003 (0.0006)	-0.0001 (0.0006)	-0.0001 (0.0010)	0.0004 (0.0009)
Gold mines	0.12 (0.33)	0.0134 (0.0130)	0.0216* (0.0131)	-0.0081 (0.0189)	0.0069 (0.0189)
N Observations		44429	44429	44429	44429
N Clusters		8446	8446	8446	8446
R2		0.23	0.23	0.25	0.27

Estimates from Ordinary least squares regressions. SE clustered at school level. Dependant variables, in the columns, are standardised Saber 11 test scores. Independent variables, in the rows, include ratios derived from the subjective income expectations, where the medium value is assigned to each response category. In Panel B, regressions include controls on students, family and counties characteristics (see data section for more details). Significance: 99% \*\*\*; 95% \*\*; 90% \*.  
Source: Authors' own calculations.

clustering specifications—instead of an individual level, we considered school and municipality level clustering of the error. Typically, standard errors are larger than in the individual-level exercise; however, almost all coefficients keep the same significance level.

In a final exercise, we considered as an outcome whether the student expected higher wages under the vocational education scenario than in high school, and a similar variable for professional education. The goal of the exercise was to simplify the assumptions required for the estimation by

reducing the problem to a binary response exercise, with the cost of a less sensitive outcome.<sup>12</sup> Most of the variables explaining the differences on the premium are the same as the ones described above: age, gender, socio-economic index, parents' education, observed wage, and wage premium in the metropolitan area, school ranking, and mean math scores.

12 The main exercise captures variation on the intensive margin of the premium, given by the different categories of the expected conditional incomes, while the main exercise only considered the extensive margin.

#### 4. Conclusions

This paper analysed the earnings and academic expectations of a random sample of 10% of the students who took the official secondary education leaving test, *Saber II*, which is required to enroll for academic or vocational studies in the country. First, we studied how these expectations correlated with individual, family, school, and municipality characteristics. Second, we considered their relation with actual test scores, which took place several weeks after the elicitation of the beliefs information.

First, we found that expected earnings in each level of education follow the trends observed in local labour markets, particularly wages for individuals 5 years older than the students tested. Contrary to Jensen's (2010) finding regarding secondary school returns in rural Dominican Republic, there is evidence that Colombian students expect a notoriously higher income once they have obtained a college degree, which is consistent with Bonilla et al.'s (2015) analysis in 120 schools in Bogotá.

Second, earnings expectations follow some patterns observed in the data, such as gender gaps and disability gaps. An interesting finding in this direction is that the gender gap is not present at the level of college education. This might be related to the fact that more girls significantly expect to enrol in college education than boys. This is conditional on school quality. However, simultaneously, during the last decade, there was a higher fraction of girls completing secondary education in Colombia (the number of girls taking the *Saber II* test is higher than boys during the last few years). Since vocational studies are, on average, more affordable for low-income individuals, it is likely that expectations toward premiums differ in each modality of post-secondary education according to parental expectations. However, as was mentioned before, parental expectations (Attanasio and Kaufman, 2009; Armand, 2015) and a decrease in demand barriers (Barrera-Osorio et al., (2008; Nopo and Hoyos, 2010) encourage women's participation and delay fertility choices (Glassman, 2015).

Third, students in private schools expect higher wages, higher test scores and higher odds of enrolling in tertiary education, despite conditioning on the socio-economic characteristics of both families and municipalities, and the observed performance of schools in prior *Saber II* tests. This result is particularly interesting, considering that test scores do not show such high-test scores premium. Such a difference in expectations indicates that enrolment in private schools is related to a family's desired academic outcome for their children, but does not guarantee a better performance once we consider their performance in prior *Saber II*. As a result, it is of primary relevance to encourage policy-makers to divulge test results, so that families can choose schools based on accurate information on their quality.

Finally, expected earnings and test scores embedded private information that explains the variation in actual test scores, even after conditioning on key characteristics. However, they only marginally improve our ability to understand the test score outcomes. As shown by Bonilla et al. (2015), it is likely that information interventions have very limited power in terms of improving tertiary education enrolment in Latin America.

#### Acknowledgments

The authors would like to acknowledge the access to databases provided by Adriana Molina, Fabio Sánchez, and Juan F. Vargas. All of the opinions expressed in the document are our own. We also thank Patrick McCartney, PhD, from Edanz Group ([www.edanzediting.com/ac](http://www.edanzediting.com/ac)) for editing a draft of this manuscript.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

The routines required to replicate the main results of this article have been made freely accessible on the Internet at: <https://github.com/androdri1/SubjectiveEarnings>

#### References

- Arcidiacono, P., Hotz, V., & Kang, S. (2010). Modelling college major choices using elicited measures of expectations and counterfactuals. Technical report, National Bureau of Economic Research. NBER Working Paper No. 15729
- Armand, A. (2015) Are parental perceived returns to schooling predicting future Schooling Decisions? Evidence from Macedonia. Navarra Center for International Development. Working Paper 2.
- Attanasio, O., & Kaufmann, K. (2009). Educational choices, subjective expectations, and credit constraints. Technical report, National Bureau of Economic Research. NBER Working Paper No. 15087
- Attanasio, O., Meghir, C., & Vera-Hernández, M. (2005). Elicitation, validation, and use of probability distributions of future income in developing countries. Unpublished working paper, University College London.
- Attanasio, O. P. (2009). Expectations and perceptions in developing countries: Their measurement and their use. *The American Economic Review* 99(2), pp. 87–92.
- Avitabile, C., & De Hoyos- Navarro, R. (2015). "The Heterogeneous effect of information on student performance: evidence from a randomized control trial in Mexico." World Bank Policy Research Working Paper, (7422).
- Barrera-Osorio, F., Bertrand, M., & Linden, L., & Perez-Calle, F. (2008) Conditional Cash Transfers in Education Design Features, Peer and Sibling Effects Evidence from a Randomized Experiment in Colombia. NBER Working Paper No. 13890

- Becker, G. S. "Human capital theory." Columbia, New York (1964).
- Bonilla, L., Bottan, N. L., & Ham, A. (2015). The Role of Information on Students' Career Choice and School Effort. Experimental Evidence from Bogotá, Colombia Retrieved from <http://publish.illinois.edu/andresham/files/2015/01/informationrct.pdf> Sept 12th/2017.
- Brunnermeier, M. K., & Parker, J. A. (2005). Optimal expectations. *American Economic Review* 95, (4), pp. 1092-1118
- Carneiro, P., Heckman, J. J. & Vytlacil, E. J. (2011). Estimating marginal returns to education. *American Economic Review* 101(6), pp. 2754–81.
- Cunha, F., Heckman, J. J. & Schennach, S. M. (2010), Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, 78 (3), pp. 883–931
- Delavande, A., Gine, X., & McKenzie, D. (2011). Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of Development Economics* 94(2), pp. 151 – 163.
- Dominitz, J. & Manski, C. (1996). Eliciting student expectations of the returns to schooling. *Journal of Human Resources*, 31 (1) pp. 1–26.
- Fryer Jr, R. G. (2013). Information and student achievement: Evidence from a cellular phone experiment. National Bureau of Economic Research. NBER Working Paper No. 19113
- García, S., Harker, A., & Cuartas, J. (2016). Building Dreams: the Impact of a Conditional Cash Transfer Program on Educational Aspirations in Colombia. Universidad de los Andes. *Documentos de Trabajo, No. 30*
- Glassman, A., Todd J., & Gaarder, M. (2015) Latin America: Cash Transfers to Support Better Household Decisions. Chapter 6. In *Performance Incentives for Global Health: Potential and Pitfalls*. Eds Rena Eichler, Ruth Levine and the Performance-Based Incentives Working Group Center for Global Development, 89-121
- Greene, W. H. (2002). The Behavior of the Fixed Effects Estimator in Nonlinear Models. NYU Working Paper No. EC-02-05
- Hansen, K. T., Heckman, J. J., & Mullen, K. J. (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics* 121(1-2), pp. 39–98.
- Hastings, J. S., Neilson, C. A., Ramirez, A., & Zimmerman, S. D. (2016). (Un) informed College and Major Choice: Evidence from Linked Survey and Administrative Data. *Economics of Education Review*.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* 47(1), pp. 153–161.
- Heckman, J. J., Lochner, L. J., & Todd, P. E. (2006). Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond, pp. 307–458. *Handbook of the Economics of Education*. Elsevier.
- Hoxby, C. & Turner, S. (2013) Expanding College Opportunities for High-Achieving, Low Income Students Stanford Institute for Economic Policy Research. SIEPR Discussion Paper No. 12-014
- ICFES (2010). Metodología de construcción del índice de nivel socioeconómico de los estudiantes -inse- y de la clasificación socioeconómica -cse- de los colegios. Technical report, Instituto Colombiano para la Evaluación de la Educación.
- Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics* 125(2), pp. 515–548.
- Klaauw, W. V. (2012). On the use of expectations data in estimating structural dynamic choice models. *Journal of Labor Economics* 30(3), pp. 521–554.
- Loyalka, P., Liu, C., Song, Y., Yi, H., Huang, X., Wei, J., ... & Rozelle, S. (2013). "Can information and counseling help students from poor rural areas go to high school? Evidence from China." *Journal of Comparative Economics*, 41(4), pp. 1012–1025.
- Manski, C. & Dominitz, J. (1996). Eliciting student expectations of the returns to schooling. *Journal of Human Resources* 31(1), pp. 1–26.
- McGuigan, M., McNally, S., & Wyness, G. (2014). Student Awareness of Costs and Benefits of Educational Decisions: Effects of an Information Campaign and Media Exposure (No. 14-19). IZA Discussion Paper No. 8596. Retrieved on Sept 20/2017 from <http://ftp.iza.org/dp8596.pdf>
- Mincer, J., & Polachek, S. (1974) "Family investments in human capital: Earnings of women." *Journal of Political Economy* 82(2), Part 2: pp. S76-S108.
- Nguyen, T. (2013). Information, role models and perceived returns to education experimental evidence from Madagascar. en *Gender Impact : the World Bank's Gender Impact Evaluation Database*. Washington DC ; World Bank.
- Ñopo H., & Hoyos, A. (2010) Evolution of Gender Wage Gaps in Latin America at the Turn of the Twentieth Century: An Addendum to "New Century, Old Disparities". IZA DP No. 5086, July
- Oreopoulos, P. & Salvanes, K. (2009). How large are returns to schooling? Hint: Money isn't everything. National Bureau of Economic Research. NBER Working Paper No. 15339,
- Pekkala Kerr, S., Pekkarinen, T., Sarvimäki, M., & Uusitalo, R. (2015). Post-secondary education and information on labor market prospects: a randomized field experiment. IZA Discussion Paper No. 9372. September
- Rao, R. (2016). Information, Heterogeneous Updating & Higher Education Decisions: Experimental Evidence from India. Retrieved on Sept 17/2017 from <http://publications.dyson.cornell.edu/grad/candidates/2016/Dyson-TanviRao-Abstract.pdf>
- Reuben, E., Wiswall, M. & Zafar, B. (2017). Preferences and biases in educational choices and labor market expectations: shrinking the black box of gender. *The Economic Journal*, 127(604), pp. 2153-2186.
- Sequeira, S., Spinnewijn, J., & Xu, G. (2016). Rewarding schooling success and perceived returns to education:

- Evidence from India. *Journal of Economic Behavior & Organization*. 131, Part A, pp. 373-392
- Stewart, M. B. (1983). On least squares estimation when the dependent variable is grouped. *The Review of Economic Studies* 50(4), pp. 737–753.
- Stinebrickner, T., & Stinebrickner, R. (2012). Learning about academic ability and the college drop-out decision. *Journal of Labor Economics*. 30 (4), pp. 707-748
- Wiswall, M., & Zafar, B. (2015). How do college students respond to public information about earnings? *Journal of Human Capital* 9, (2), pp. 117-169.
- Zafar, B. (2011). How do college students form expectations? *Journal of Labor Economics* 29(2), pp. 301–348.
- Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources* 48(3), pp. 545–595.

## Appendix Data Restrictions

Sample restrictions Variables	Obs	Percent
Saber 11: Verbal Scores of 0 or above 100	0	0.00%
Saber 11: Maths Scores of 0 or above 100	26	0.04%
Test-taker aged above 30	1890	3.25%
Night journey school	4974	8.54%
Weekends journey school	3168	5.44%
<b>Sample size</b>		
Saber 11 dataset	575223	
Expectations Sample	58224	100.00%
Sample after restrictions (% original sample)	49361	84.78%
Missing on controls Variables	Missing	Percent
Father Ed: Less than High School	11	0.02%
Father Ed: Vocational Education	11	0.02%
Father Ed: Professional Education	11	0.02%
Mother Ed: Less than High School	11	0.02%
Mother Ed: Vocational Education	11	0.02%
Mother Ed: Professional Education	11	0.02%
Saber 11 School Ranking: High and above	21397	36.75%
Saber 11: Std Mean Maths Scores 2012	23199	39.84%
Saber 11: Std Mean Verbal Scores 2012	23199	39.84%
Non Mixed Gender	18158	31.19%
Wage premium 16 to 11 years of educ.	12258	21.05%
Avg. wage 11 years of educ.	12258	21.05%
MD. Poverty rate	656	1.13%
Urban-Rural Ratio	637	1.09%
Homicide rate 2009	637	1.09%

Source: Own calculations based on a 10% student sample from Saber 11 2013-II.