The Impact of Child Labor on Student Enrollment, Effort and Achievement: Evidence from Mexico

Gabrielle E. Vasey*

Department of Economics, University of Pennsylvania

Latest Version Here

This Version: November 9, 2020

Abstract

When school-age children work, their education competes for their time and effort, which may lead to lower educational attainment and academic achievement. This paper develops and estimates a model of student achievement in Mexico, in which students make decisions on school enrollment, study effort and labor supply, taking into account locally available schooling options and wages. All of these decisions can affect their academic achievement in math and Spanish, which is modeled using a value-added framework. The model is a random utility model over discrete school-work alternatives, where study effort is determined as the outcome of an optimization problem under each of these alternatives. The model is estimated using a large administrative test score database on Mexican 6th grade students combined with survey data on students, parents and schools, geocode data on school locations, and wage data from the Mexican census. The empirical results show that if students were prohibited from working while in school, the national dropout rate would increase by approximately 20%, while achievement would increase in both math and Spanish. Expanding the conditional cash transfer, either in terms of the magnitude of the cash benefits or the coverage, in conjunction with prohibiting working while in school is an operational policy that would greatly reduce dropout while maintaining the gains in achievement.

*gvasey@sas.upenn.edu, 133 South 36th St., Philadelphia, US

I am very grateful to my advisors Jere Behrman, Holger Sieg, Petra Todd, and Arthur van Benthem for their help and guidance. I would further like to thank my friends and colleagues at the department for their encouragement and support, as well as participants of the Empirical Micro Student Seminar for helpful comments.

1 Introduction

When children participate in the labor force, it is often at the expense of their education. Globally, the International Labour Organization estimated that 265 million children were working in 2013. The trade-off between working, with the benefits of receiving a wage or helping family, and attending school, in the hope of increasing future wages, is one that many children and families face worldwide. Many children who attend school also work part time and face another choice with respect to the amount of time and effort to dedicate to studying compared to working. Family socioeconomic status, school availability, school quality, ability and earnings opportunities all influence children's time allocation decisions and their resulting academic achievement and attainment.

This paper explores the determinants of child labor, school enrollment and academic achievement in Mexico. I consider children who have graduated from primary school (Grade 6) and who should be enrolling in middle school (Grade 7). Mexican Basic Education, defined as Grades 1 through 9, is compulsory and labor of minors under the age of 14 is legally prohibited. However, in the 2010 Census, 7.9% of children aged 12 and 13 report not enrolling in school. A nationally representative survey in 2009 found that 25.7% of Grade 7 students who are in school report working at least one day a week. Many developing countries around the world face similar struggles to keep children in school and out of the labor force.

To study the determinants of children's time allocation decisions, I develop and estimate a model of school and labor participation decisions with endogenous school effort choices. In my model, individuals who finish primary school have a choice set of middle schools available. The choice set is determined using data on school locations and prior-year school attendance patterns. The middle schools are treated as differentiated products that vary in terms of school infrastructure and principal characteristics such as experience, as well as the type of school curriculum. The choice of school affects a student's utility directly, as well as their achievement production function and marginal cost of effort. Effort is costly, and the marginal cost of effort depends on student characteristics and on whether the student is working. Wage offers vary by student demographics and by primary school location and there are separate wage offers for working full time and working while enrolled in school.

I use the estimated model to evaluate how education and work-related policy changes would affect school enrollment, academic achievement, and children's labor-force participation rate. First, I use the model to simulate the effects of a policy that removes the labor option for children who are enrolled in school. These estimates provide insight into what fraction of students only attend school if they can also work and how much achievement would increase if students did not divide their time between work and school. The second and third counterfactuals consider policies that would work in conjunction with the first, with the goal of reducing the drop out rate. The second counterfactual considers prohibiting all child labor and the third considers expanding the conditional cash transfer for school attendance, both in terms of benefits amounts and program coverage.

To estimate my model, I combine several data sources: administrative data on nationwide standardized tests in math and Spanish, survey data from students, parents and principals, geocode data on school locations, and Mexican census data on local labor market wages and hours worked. The administrative data include information on which students are beneficiaries of Prospera, the conditional cash transfer program.

In the model, students decide whether to attend school, and if they attend, also decide what type of school to attend and how much effort to dedicate to their studies. The marginal cost of effort varies by age, gender, parental education, family income, distance traveled to school, and working status. I use the model's first-order conditions to solve for an optimal effort level that is specific to the type of school. Schools differ in the demands they place on students as well as the required travel time. The data provide five measures of self-reported effort that are used within a factor model to determine a single effort index.

Using the administrative test score data, I estimate value-added achievement production functions that incorporate student's effort choices. Separate equations are estimated for math and Spanish test scores. The specification includes lagged test scores, along with student demographics, school characteristics, and efforts. The unobserved components of the student's math and Spanish test scores are allowed to be correlated.

In the model, students enrolled in school may choose to work part time, and students who do not enroll in school are assumed to work full time. The model incorporates a wage offer equation that represents the utility children receive if they do not attend school. To capture the monetary value of working, I estimate hourly wages and the numbers of hours worked for boys and girls using Census data and a Heckman (1976) selection model. With these estimates, I construct wages that vary by geographic location, urban area, age, and parental education.

The model is a discrete-continuous choice model with partially latent continuous choice variables (Dubin and McFadden, 1984). Specifically, it is a random utility model over discrete school-work alternatives, where study effort is determined as the outcome of an optimization problem under each of the school-work alternatives. I estimate the model via Maximum Likelihood, where the probability can be decomposed into three conditional probabilities, which each have a closed-form solution.

I find that traveling to a middle school is costly and that students value distance education schools (Telesecondaries) less than the other two school types (General and Technical). Students value schools with high average expected test scores, however the amount of weight they put on that component does not depend on their parents education levels or if they are conditional cash transfer beneficiaries. Effort is costly to students, especially when working, but less so for female students and for students with higher lagged test scores. Students are estimated to dislike working while in school overall. Effort is estimated to be an important input into both math and Spanish test score production functions.

The results of the counterfactual analysis show that almost 10% of students who are working while enrolled in school would drop out if they were unable to combine work and school. This increases the national dropout rate by almost 20%. For students who remain enrolled, their effort increases by approximately 3% of a standard deviation, resulting in increases in their math and Spanish scores by an average of 3% of a standard deviation. Prohibiting all child labor results in a dropout rate lower than under the benchmark model. However, a similarly low dropout rate can be achieved by either increasing the conditional cash transfer amounts, or expanding the set of families who receive the conditional cash transfer to include more of those with low monthly incomes.

Recently, there have been several papers estimating discrete choice models to estimate models of school choice, where schools with differing characteristics are treated as differentiated products (Epple, Jha and Sieg, 2018; Neilson, 2013). These models are similar to mine in that they include school characteristics and a student achievement production function, and the authors use the model to evaluate how policy changes impact school choices. I extend these frameworks by allowing for dropping out of school and part-time or full-time work. I also incorporate students' decisions of how much effort to devote to their studies. These extensions are needed to make the model relevant to developing country contexts.

A large portion of the literature examining the relationship between child labor and education considers how policies, such as conditional cash transfers, affect school enrollment and child labor. Dynamic models have been used to evaluate the long-term effect of such policies, however none thus far has incorporated test score production functions, time allocation decisions, and decisions about what type of school to attend (Attanasio, Meghir and Santiago, 2011; Todd and Wolpin, 2006). There also exist some static occupational choice models that include the options of dropping out, enrolling and working part time, or only enrolling. However, these models also do not examine academic achievement or how working part time affects a child's ability to study (Bourguignon, Ferreira and Leite, 2003; Leite, Narayan and Skoufias, 2015). Finally, there are some recent papers that consider the impact of labor on achievement, without incorporating school choice and enrollment. Keane, Krutikova and Neal (2018) consider many possible uses of time for students, and find that working is only harmful to achievement if it is taking away from study time.

Although there is a substantial literature in the education economics field studying teacher effort, how it affects student achievement and how it is influenced by incentive pay, there is relatively little focus on student effort, which is an important input in academic achievement. A study using an instrumental variables approach found that school attendance has a positive causal impact on achievement for elementary- and middle-school students (Gottfried, 2010). A causal relationship between study time and grades has also been found for college students (Stinebrickner and Stinebrickner, 2008). There are very few papers that model student effort in a structural way, and estimate how it affects learning. Todd and Wolpin (2018) develop and estimate a strategic model of student and teacher efforts within a classroom setting.

The literature on CCT programs, and specifically on the Prospera program, is extensive. The program began in 1997, and since then over 100 papers have been written about it (Parker and Todd, 2017). The majority of these papers use the experimental data gathered during the first two years of the program. There is a consensus in the literature that Prospera increases enrollment in school for students in junior and senior high school (Attanasio, Meghir and Santiago, 2011; Behrman, Sengupta and Todd, 2005; Dubois, De Janvry and Sadoulet, 2012; Schultz, 2004). However, studies focused on student enrollment and grade progression and not on student achievement, with the exception of two recent working papers (Acevedo, Ortega and Székely, 2019; Behrman, Parker and Todd, 2019)[1]. Finally, there are a few studies using experimental data to estimate the impact of conditional cash transfers on child labor decisions. For example, Yap, Sedlacek and Orazem (2009) find that the PETI

[1]These papers use matching and regression-based treatment effect estimators.

program in Brazil increased academic performance and decreased child labor for beneficiary households.

The paper proceeds as follows. Section 2 describes the model of discrete school-work alternatives with endogenous effort choice. Section 3 describes the dataset and summary statistics for the variables of interest. Section 4 provides empirical evidence on the relationships between working, effort and school achievement. Section 5 describes the estimation strategy and Section 6 discusses the results from the estimation. Section 7 discusses the policy implications and Section 8 concludes.

2 Model

The model captures the different choices that students make as they progress from primary school (Grade 6) to middle school (Grade 7). The first choice is what school, if any, they wish to attend. Based on the location of the primary school that student i attended (P_i), the student will have a choice set of available middle schools, $\mathbb{S}_{P_i}$. Middle schools are categorized into three types: General, Technical (vocational) and Telesecondaries (distance education). Students also make a labor choice. If the student chooses not to enroll in school, it is assumed that they work full time. Students who choose to enroll in school may choose between working part time or focusing only on their studies. Students receive wage offers that depend on their age, gender, parental education, location and whether they are enrolled in school. Finally, students who enroll in school make an effort choice. Effort is costly, however it is an input into the achievement production function and students' utility depends on achievement.

Each student who finished Grade 6 enters the model with a set of initial conditions. These include their gender, their age, their lagged test scores and if they are a beneficiary of Prospera, the conditional cash transfer program. Also included are permanent family characteristics including the number of siblings, the parental education levels, the monthly family income, and some information about the household, such as if they own a computer. Finally, the geographic location of the primary school is included, which gives information on whether the neighbourhood is rural or urban, and also identifies the choice set of middle schools.

2.1 Student Utility

Students in the model are 12 years old on average, and therefore it is plausible that they are making their schooling choice along with their family. Families care about student achievement, monetary compensation coming from Prospera or wages, the type of school the student attends, the cost of traveling to school, and the cost of effort. Effort may be more costly if the student has other demands on their time, such as a part time job, or if they have lower lagged test scores. The utility of student i attending school j is given by

$$U_{ijL}(e_{ijL}) = \underbrace{CCT_i + \mathbb{1}\{L = PT\}w_i^{PT}}_{\text{Monetary Compensation}} + \underbrace{\alpha_1 d_{P_ij} + \alpha_2 d_{P_ij}^2}_{\text{Distance Traveled}} +$$
$$\underbrace{(\alpha_3 + \alpha_4 PEduc_i + \alpha_5 \mathbb{1}\{CCT > 0\}) \left(\hat{A}_{ij}^{7,S}(e_{ijL}) + \hat{A}_{ij}^{7,M}(e_{ijL})\right)}_{\text{Achievement}} +$$
$$\underbrace{\alpha_6 + \alpha_7 PEduc_i + \sum_{k \in Type} \beta_k \mathbb{1}\{Type_j = k\}}_{\text{School Types}} + \underbrace{\alpha_8 \mathbb{1}\{L = PT\}}_{\text{Working}} +$$
$$\underbrace{(\alpha_{i,9} + \mathbb{1}\{L = PT\}\alpha_{10})\, e_{ijL} + \alpha_{11} e_{ijL}^2}_{\text{Effort}} + \nu_{ijL}$$

The monetary compensation includes the conditional cash transfer CCT_i, which student i receives if they are a Prospera beneficiary, as well as a part-time wage w_i^{PT}, which they receive if they choose to work part time. The coefficient on the monetary component is constrained to one, so that the units of the remaining utility coefficients are in terms of money (pesos). The distance between student i's primary school P_i, and middle school j is given by d_{P_ij}. Achievement in Spanish and math, $\hat{A}_{ij}^{7,S}(e_{ijL})$ and $\hat{A}_{ij}^{7,M}(e_{ijL})$, depend on student characteristics, middle-school characteristics, and students' effort choices e_{ijL}. Students may care differently about their scores depending on their parent's education, $PEduc_i$ and if they are a conditional cash transfer beneficiary. To capture parental education, $PEduc_i$ is equal to one if both parents have at least a middle-school education. Students receive a benefit from enrolling in school, which is captured by α_6, and this benefit may vary depending on parental education. $Type_j$ is school j's type, and can be one of Telesecondary, Technical or General. Students potential distaste for working while in school is captured by the coefficient α_8.

The random coefficient $\alpha_{i,5}$ captures heterogeneity in the marginal cost of effort across students. The coefficient can be broken down into a component that is constant across students, a component that varies with student characteristics, and a random unobserved component,

$$\alpha_{i,9} = \alpha_9 + \lambda X_i + \eta_i$$

where $\eta_i \sim \mathcal{N}(0, \sigma_\eta^2)$. Student characteristics contained in X_i include the students' gender, their parental education, and their lagged test scores.

If students choose the outside option, they are choosing to drop out of school after 6th grade. It is assumed that they work full time, and receive a full time wage w_i^{FT}.

$$U_{i0} = w_i^{FT} + \nu_{i0}$$

The error terms are assumed to be iid type I extreme value, so the overall framework is a mixed logit model. The wages, w_i^{PT} and w_i^{FT} are estimated using Mexican Census data as described below.

Student i's choice set of middle schools, $\mathbb{S}_{P_i}$, is comprised of all middle schools within a certain distance of their primary school, P_i. This distance is computed by considering how far students have historically traveled from this primary school. Because of this, some choice sets cover smaller areas than others. Each school in the choice set is defined by the distance between it and student i's primary school, d_{P_ij}, and the type of school it is, $Type_j$. Other school-level variables from the principal survey that I include in the analysis relate to infrastructure and principal and teacher quality.

2.2 Wage Offers

Each student receives a full-time and a part-time wage offer. If they accept the full time wage, they are not able to enroll in school. They can also choose to not accept either offer and only enroll in school. Potential hourly wages for children are imputed using Census data. Wages are allowed to depend on age, gender, school attendance, parental education, and geographic location (either urban/rural and municipality). To account for non-random selection into working, a Heckman selection model was estimated. Variables representing family socioeconomic levels, such as family income and home infrastructure are used as instruments that affect selection into working, but do not affect the wage offers directly. Regressions are estimated separately for girls and boys. For details on the wage estimation and parameter estimates, see Appendix A.1.

$$w_{igj} = \gamma_0 + \underbrace{\gamma_1 a_i}_{\text{Age}} + \underbrace{\gamma_2 \mathbb{1}\{j \neq 0\}}_{\text{Not enrolled}} + \gamma_3 a_i \times \mathbb{1}\{j \neq 0\} + \underbrace{\gamma_4 MomEduc_i + \gamma_5 DadEduc_i}_{\text{Parental education}}$$
$$+ \gamma_6 a_i \times MomEduc_i + \gamma_7 a_i \times DadEduc_i + \underbrace{Geo_g}_{\text{Municipality FE}} + \nu_{igj}$$

2.3 Expected Test Scores

For students who choose to enroll in school, their test scores are generated by a value-added production function. The student inputs to the production function include lagged test scores, student characteristics (including age, gender, and family characteristics) and their effort choices. School inputs, Z_j, include the type of school, principal education and experience, if the school has internet, if the school teaching materials are sufficient, and how the principal rates the teachers.

$$\hat{A}_{ij}^{7,T}(e_{ijL}) = \delta_0^T + \underbrace{\delta_1^T A_i^{6,M} + \delta_2^T A_i^{6,S}}_{\text{Lagged Scores}} + \underbrace{\delta_3^T e_{ijL}}_{\text{Effort}} + \underbrace{\delta_4^T X_i}_{\text{Student chara.}} + \underbrace{\delta_5^T Z_j}_{\text{School chara.}} + \delta_6^T e_{ijL} Z_j + \xi_{ije}^T$$
$$\text{for } T \in \{S, M\} \quad (1)$$

The value-added equation is estimated separately for math and Spanish test scores. For each student, the math and Spanish residuals are allowed to be correlated. Students are assumed to not know the error terms when making their school choices. Working does not directly affect achievement. However, working makes study effort more costly. The benefits of effort may vary by school type.

2.4 Maximization Problem

Student i solves the following maximization problem for their optimal level of effort e_{ijL}^* for each possible school j and labor option L in their choice set:

$$\begin{aligned} e_{ijL}^* = \underset{e_{ijL}}{\text{argmax}} \quad & U_{ijL}(e_{ijL}, \hat{A}_{ij}^{7,S}(e_{ijL}), \hat{A}_{ij}^{7,M}(e_{ijL}); X_i, Z_j, w_i^{PT}, w_i^{FT}) \\ \text{s.t.} \quad & \hat{A}_{ije}^{7,S} = f_S(A_i^{6,M}, A_i^{6,S}, e_{ijL}; X_i, Z_j) \\ & \hat{A}_{ije}^{7,M} = f_M(A_i^{6,M}, A_i^{6,S}, e_{ijL}; X_i, Z_j) \end{aligned}$$

The first-order equation of the above maximization problem yields the following expression for optimal effort:

$$e^*_{ijL} = \frac{-\big((\alpha_3 + \alpha_4 PEduc_i + \alpha_5 \mathbb{1}\{CCT > 0\})(\delta^S_3 + \delta^S_6 Z_j + \delta^M_3 + \delta^M_6 Z_j) + \alpha_{i,9} + \mathbb{1}\{L = PT\}\alpha_{10}\big)}{2\alpha_{11}} \tag{2}$$

The parameter $\alpha_{i,9}$ is a function of the student characteristics, X_i, and the random shock, η_i. The optimal effort therefore depends on student characteristics, school characteristics, labor status, and an idiosyncratic preference shock.

Define the dummy variable $D_{ijL} = 1$ if student i chooses school j and labor option L. Student i then solves the following maximization problem, given their solutions for optimal effort e^*_{ijL} and the expected achievement that the optimal effort implies ($\hat{A}^{7,S}_{ije^*}$ and $\hat{A}^{7,M}_{ije^*}$).

$$\max_{j,L} \quad \sum_{j=1}^{J_i} \sum_{L \in \{0,PT,FT\}} D_{i,j,L} \times U_{ijL}(e^*_{ijL}, \hat{A}^{7,T}_{ij}(e^*_{ijL}), \hat{A}^{7,T}_{ij}(e^*_{ijL}); X_i, Z_j, w^{PT}_i, w^{FT}_i)$$

3 Data

To carry out this research, I use a newly available merged dataset. This dataset is comprised of three separate components, that come from two sources. The first component is the Evaluación National de Logro Académico en Centros Escolares or ENLACE test scores. These tests were administered at the end of the school year to gather information on students' achievement in math and Spanish. They were given to students every year between the 2006/2007 school year and the 2013/2014 school year. The Mexican Secretariat of Public Education (SEP) was in charge of administering the test. The second component comes from the same source as the ENLACE test scores, and can be easily merged with the test score data. Every year a group of schools was randomly selected and all students enrolled in those schools were given a questionnaire. These data have recently been used for impact evaluation studies of the Prospera program (Acevedo et al., 2019; Behrman et al., 2019). The final component of the data set is comprised of a list of all schools in Mexico, and can be merged with the above data to provide the geographical location of the schools.

The test score data provides important information regarding student achievement, however whether a student took the test or not may not always be an accurate method of recording school attendance. It is possible that a student who is enrolled and attending school does not write the ENLACE test for several reasons. To ensure that these students are recorded as enrolled, even without a test score, I merge the National Student Registry

(Registro Nacional de Alumnos) with the test score data. This provides information on enrollment for all students in the country.

Finally, the model requires data on wages, which are not recorded in any of the previously mentioned sources. The 2010 Census is used to access information on children between the ages of 12 and 20, and their working status and wages. The Census also contains other personal information on the students such as their age, gender, school attendance history, parental education, living situation, and the municipality in which they reside.

Combining all of the data from the above sources, yields an incredibly rich representative sample of students across Mexico. For each student, we have their national standardized test scores, their school IDs (with associated school information), individual demographics, household demographics (including CCT status), and the average municipal wage conditional on age, gender, family background and school attendance.

3.1 Estimation Sample

The analysis in this paper focuses on students in Grade 6 in 2008 who progress to Grade 7 in 2009. The sample can be divided into two groups: those who enrolled in school in Grade 7, and those who dropped out of school after Grade 6.[2] There are 229,199 students enrolled in Grade 6 in 2008 for whom I have survey answers from themselves and their parents. Of these students, 17,195, or 7.5% do not appear in either the ENLACE data or the Roster data in any of the next four years. These students are assumed to have dropped out.

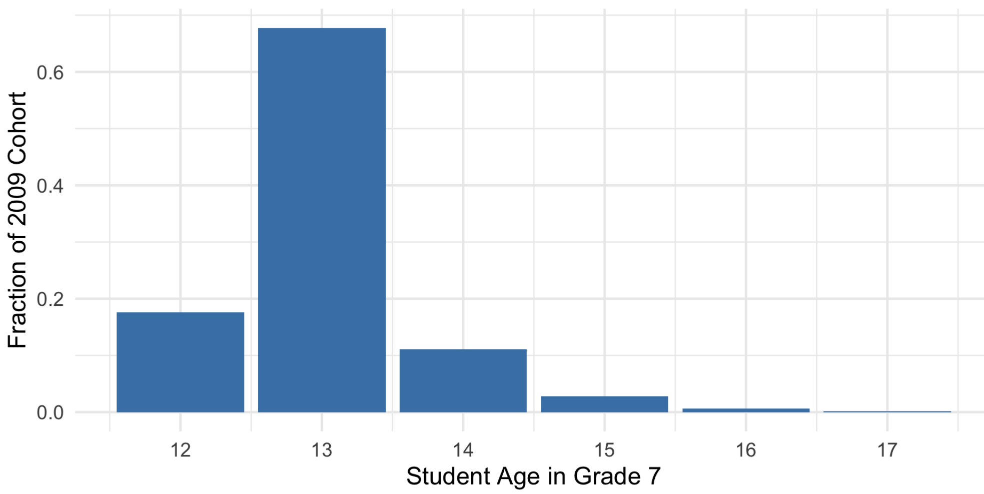
In Grade 7 in 2009, there are 107,898 students for whom we have survey answers from themselves and their parents.[3] The mean age of the students in Grade 7 is 13, and a bar graph showing the distribution of ages is shown in Figure 1. The sample is approximately equal in terms of gender, as 49.9% of the students are female. 26.1% of students are beneficiaries of the conditional cash transfer Prospera.

There are three states that are not included in the analysis. The states of Guerrero, Michoacán, and Oaxaca had many schools for which there were no ENLACE scores submitted. To prevent bias in the analysis, students in these states were not included.

[2]See Appendix A.2 for more details on how the sample for estimating the discrete choice model is constructed.

[3]Each year a different sample of schools is given the questionnaire, so the majority of these students are not in the sample of Grade 6 students from the previous year. Sample size also changes from year to year.

Figure 1: The distribution of ages of students in Grade 7 in 2009 in the estimation sample.



3.2 Summary Statistics

Standardized test scores in math and Spanish are used as a measure of student achievement. The test scores are standardized to have a mean of 500 and a standard deviation of 100. All students write the test in Grade 6 and Grade 7, so it is possible to see how they change relative to students in the same grade from their baseline results. Table 1 shows the mean and the standard deviation of Grade 6 and 7 test scores in math and Spanish. To compute these statistics, the cohort of Grade 7 students was used. This means that the students who dropped out are not included in the means, and because of this, the mean Grade 6 scores are greater than 500.

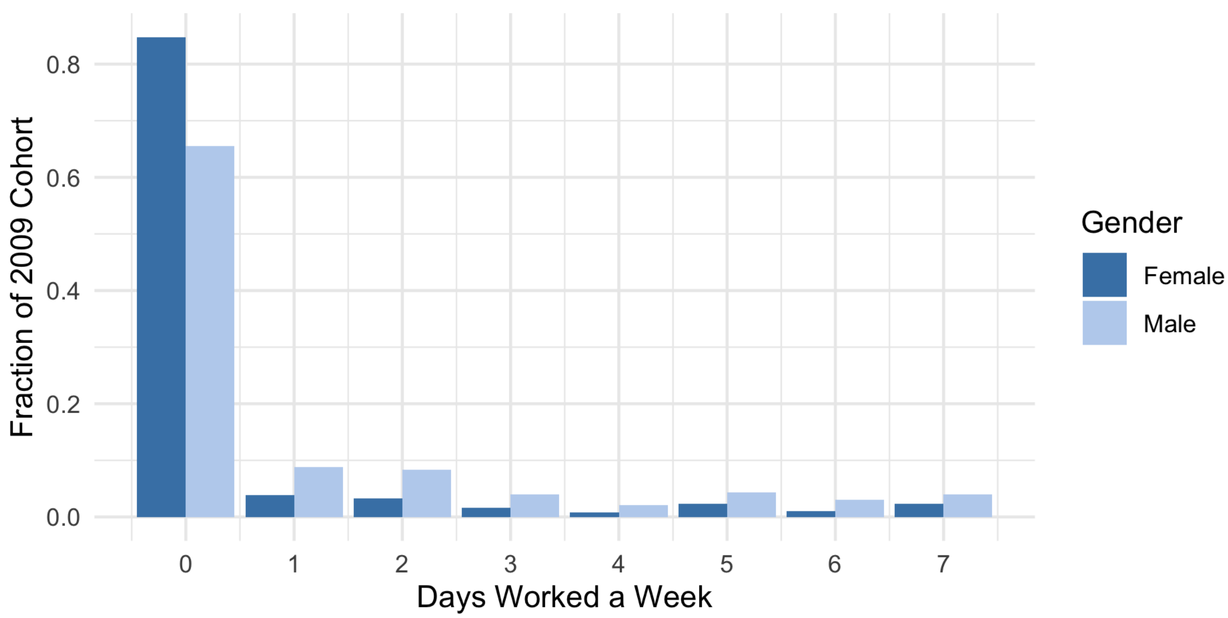
Table 1: Summary statistics for the test scores of the cohort of students in Grade 7 in 2009. The distribution of test scores are approximately normal, as shown in Appendix A.3.

	Mean	Standard Deviation
Grade 6 Math	531.2	122.5
Grade 6 Spanish	524.8	108.5
Grade 7 Math	501.0	101.5
Grade 7 Spanish	499.2	101.3

A second important variable in the model is the labor choice of each child. In the student survey, there is a question that asks: “On average, how many days a week do you work?”.

Boys work more than girls as shown in Figure 2. The mean number of days a week worked for the whole sample is 0.83. However for children 13 and younger the mean is 0.80, and for children 14 and older the mean is 1.68, so the older children are working substantially more than the younger children.

Figure 2: The distribution of the number of days worked per week, divided by gender, for students in Grade 7 in 2009 in the estimation sample.



An important contribution of this paper is that it uses data from self-reported student-effort measures. The five questions that are used as measures of effort are:

1. On average, how many hours a day do you spend studying or doing homework outside of school hours? Options: 0, 1, 2, 3, or 4 hours.

2. How often do you pay attention in your classes at school? Options: never, almost never, sometimes, almost always, always.

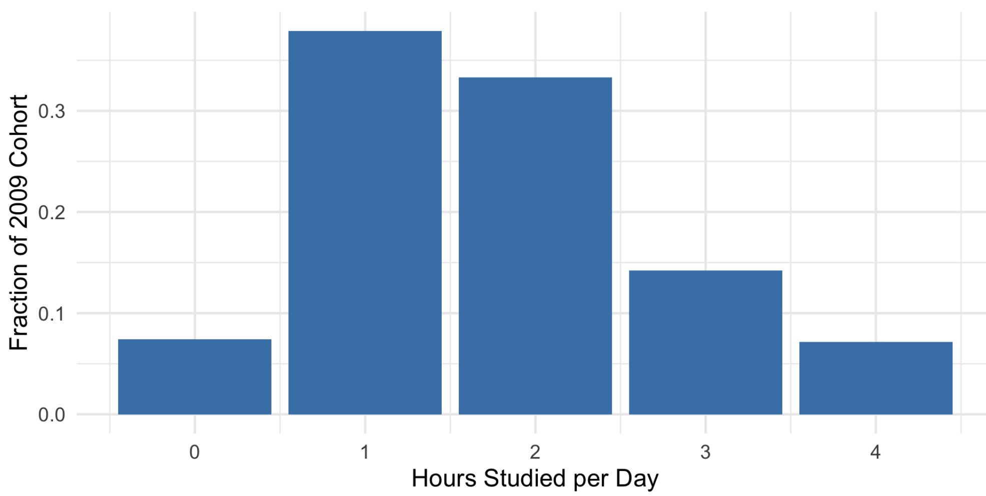
3. How often do you participate in your classes at school? Options: never, almost never, sometimes, almost always, always.

4. How often do you miss school? Options: never, almost never, sometimes, almost always, always.

5. How often do you skip your classes when you're at school? Options: never, almost never, sometimes, almost always, always.

Figure 3 shows the distribution of responses to the average number of hours studied per day. The most common response is 1 hour per day, however over 60% of students chose another response. The distribution of responses to the positive effort questions are shown in Figure 4. Most students answered "Sometimes" when asked how often they participated in class, with less students in both tail. The responses for paying attention in class follow a different pattern, with the majority of students answered "Almost Always" or "Always". Finally the distribution of responses to the negative effort questions are shown in Figure 5. The majority of students respond that they never miss school, however approximately 5% of students answer "Always" to this question. Skipping school seems to be more common, with a fair number of students answering "Almost Never" and "Sometimes".

Figure 3: The distribution of the average number of hours studied a day for students in Grade 7 in 2009 in the estimation sample.



3.3 School Types

There are four different types of middle schools in Mexico: General, Technical, Telesecondaries, and Private. Technical middle schools have a focus on vocational studies. Telesecondaries, which are wide spread and well established in Mexico, are predominately located in rural areas and offer instruction through video sessions at local centers. The purpose of these schools are to provide access to education for students in rural areas without having to incur the cost of hiring teachers specializing in each subject. Private schools are almost

Figure 4: The distribution of the positive effort measures for students in Grade 7 in 2009 in the estimation sample. The first question is how often the student pays attention in class, and the second question is how often they participate in class.

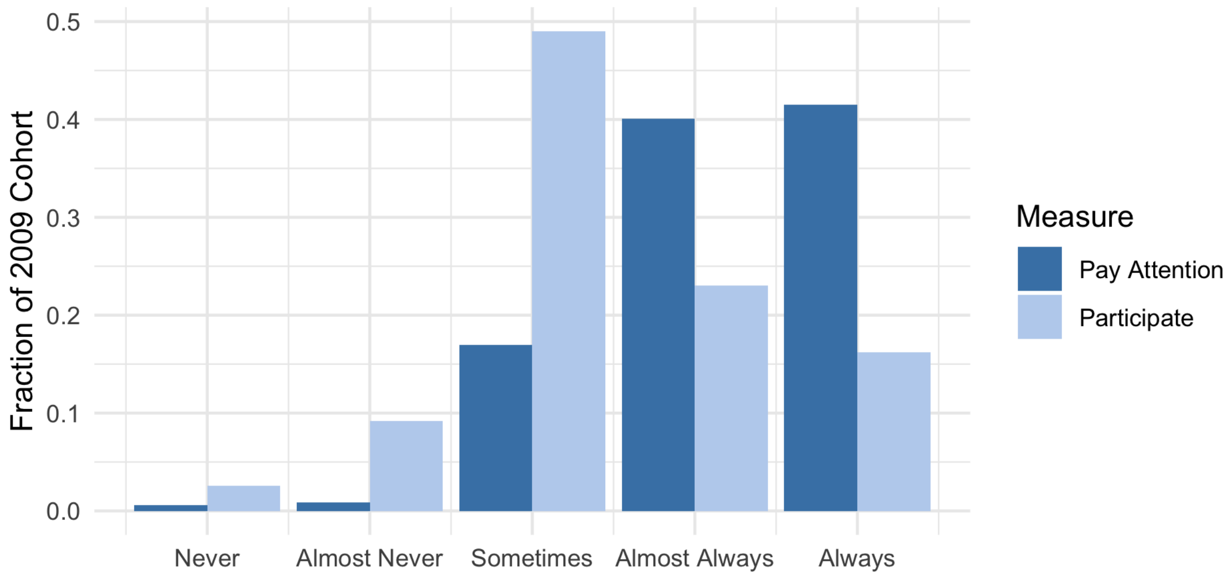
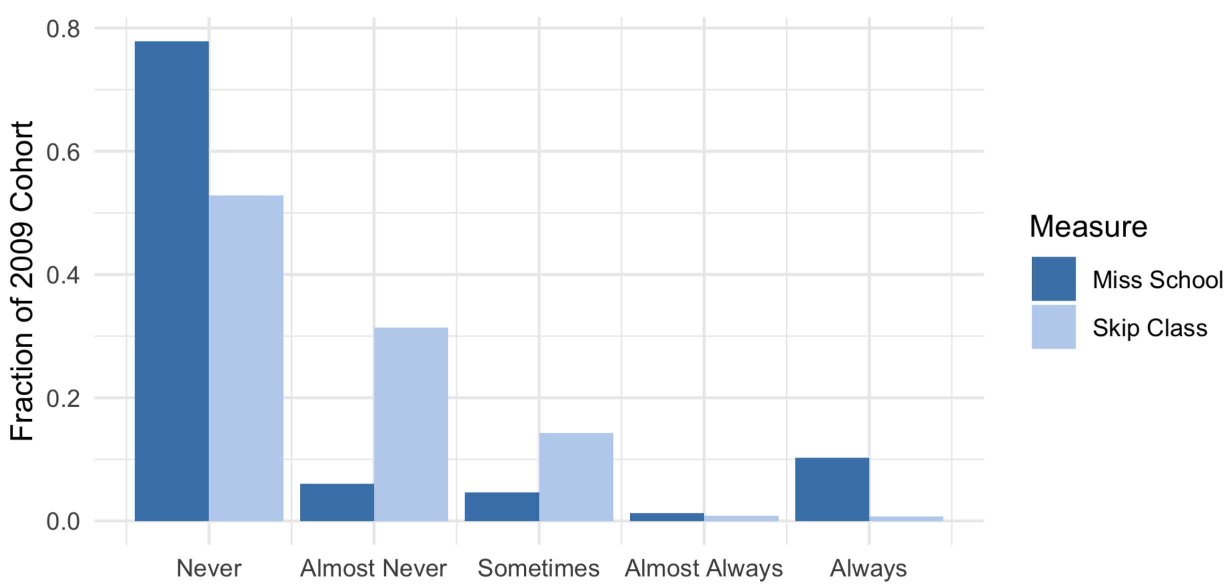


Figure 5: The distribution of the negative effort measures for students in Grade 7 in 2009 in the estimation sample. The first question is how often the student does not attend school, and the second question is how often they do not attend class while at school.



exclusively in urban areas, and have tuition payments. Unfortunately, I was not able to collect information on school tuition, so students attending private schools are not included in the estimation of the model.

Table 2 contains summary statistics for the four different types of schools in Mexico. From the table it is apparent that there are many small Telesecondary schools, predominately in rural areas. The class size of Telesecondary schools is also noticeably smaller than both General and Technical schools. Although all schools have a fairly equal amount of female and male students, the proportion of students who are beneficiaries of the conditional cash transfer differs drastically by school type. The majority of students enrolled in a Telesecondary school are beneficiaries, while less than 15% of those in General schools are. Finally, by dropping all Private schools, only 8% of students are removed from the sample.

Table 2: Summary statistics for the four types of middle schools. All data on Grade 7 students in 2009 is used to create this table.

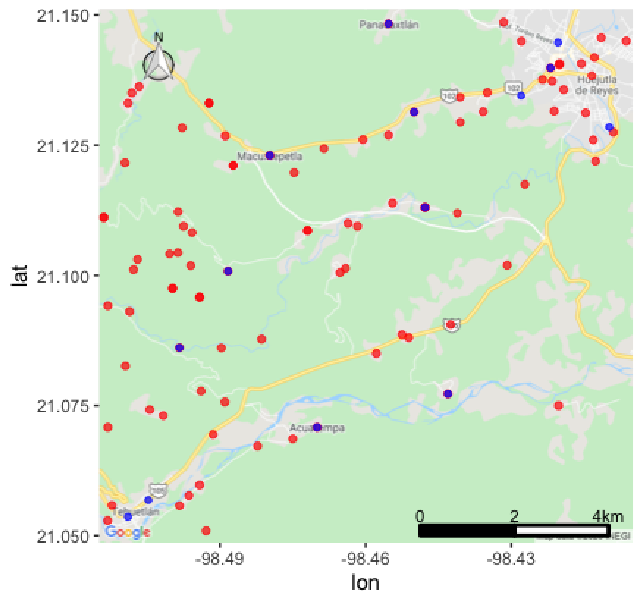
	General	Technical	Telesecondary	Private
Number of Schools	5,820	2,857	15,974	3,866
Proportion of Cohort	0.448	0.272	0.196	0.083
Proportion Female	0.500	0.497	0.492	0.504
Proportion CCT	0.166	0.195	0.642	0.008
Proportion Rural	0.146	0.230	0.873	0.021
Mean Class Size	32.3	33.9	16.6	23.6
Mean School Cohort Size	137	170	22	39

3.4 Defining Choice Sets

The location of each school in the data set is known. With these locations, it is possible to compute the distance between a student's primary school and middle school, and analyze how far students are traveling. Further, it is possible to see what other options were available within a certain distance. Examining the data, it is apparent that middle schools are much more sparse than primary schools, especially in rural regions of Mexico. Figure 6 shows the geographic distribution of primary and middle schools in a region in Mexico. Although there is a small city in the top right corner, the remainder of area covered by the map is rural. Depending on which primary school a student attended, there may be a middle school at the same location, or the nearest one may be several kilometers away.

Unfortunately the home address of students is not included in the data. Given the broad coverage of primary schools, I am assuming that students attend a primary school close to

Figure 6: Map of all primary schools (red) and middle schools (blue) in a rural region



their home, and therefore their primary school address is an adequate proxy for their home address. To calculate distance, a straight line is measured between the primary school and the middle school, as shown in Figure 7. It is also possible to calculate distance using roads and paths on Google Maps, but this does not capture many of the rural pathways.

For the estimation, I have to define which middle schools each student considers when making their school choice. To do this, I create a circle around the primary school and consider all middle schools within the circle to be in the choice set, as shown in Figure 8. However, choosing the same radius for all primary schools would not account for regional topography or the local availability of schools. Therefore, each primary school has a custom radius that is computed by analyzing how far students from that primary school traveled on average to attend middle school in previous years.[4]

[4]Distances are capped at 15km to get rid of outliers and students who moved. Students who changed state are also removed from the estimation sample.

Figure 7: Map of a primary school and several middle schools demonstrating how distance is calculated

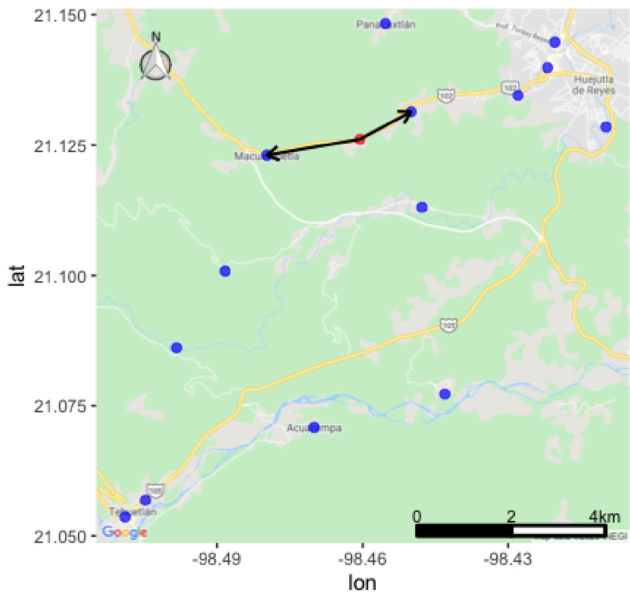
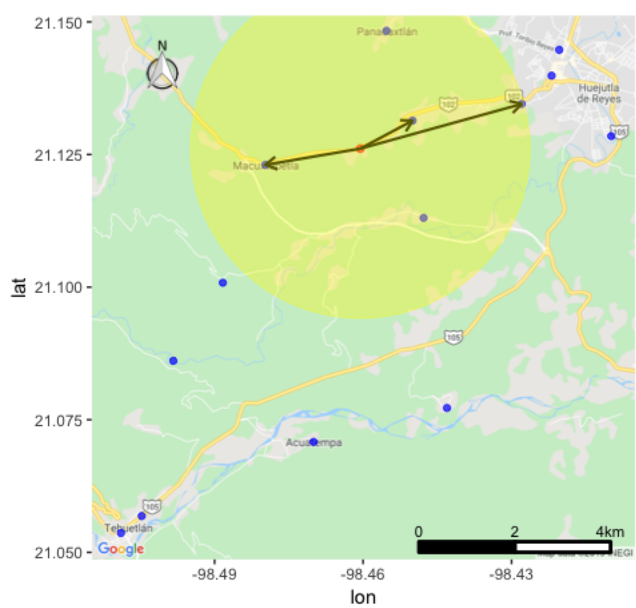


Figure 8: Map of a primary school and the middle schools included in its choice set



4 Data Description

The main mechanism through which working part time can affect student enrollment and achievement in my model is effort. I assume that expending effort on school is costly, and that this cost increases when a student is also working, as they have less time for studying. Using the effort questions from the survey questionnaires, I show correlations in the data to support these assumptions. In the following tables, the first effort measure, the number of hours studied per day, is treated as a continuous measure, and the other four measures are treated as ordinal discrete variables.

Table 3 shows that students who put in more study hours are also the students who have higher test scores. Test scores have been transformed so that they have a standard deviation of 1. The table can then be interpreted as showing an increase of studying 1 hour per day is correlated with an increase in test scores of 10% of a standard deviation. If student characteristics and lagged test scores are included in the regression, the effect of study hours decreases, but it is still positive and significant.

Table 3: Correlation between hours studied and test scores

	Dependent variable:			
	Spanish_7		Math_7	
	(1)	(2)	(3)	(4)
StudyHours	0.116***	0.045***	0.105***	0.048***
	(0.003)	(0.003)	(0.003)	(0.003)
Age		−0.042***		−0.038***
		(0.004)		(0.004)
Female		0.236***		−0.066***
		(0.005)		(0.005)
Math_6		0.135***		0.469***
		(0.004)		(0.004)
Spanish_6		0.510***		0.216***
		(0.004)		(0.004)
Constant	4.742***	1.989***	4.786***	1.908***
	(0.007)	(0.052)	(0.007)	(0.052)
Observations	82,580	82,580	82,580	82,580
R^2	0.014	0.426	0.012	0.431
Note:			*p<0.1; **p<0.05; ***p<0.01	

Table 4 and Table 5 show the correlations between the other four measures of effort and Spanish and math test scores. There are five possible answers to the questions (Never, Almost Never, Sometimes, Almost Always, and Always), and for the two tables the base level is assumed to be Never. The first two columns of both tables are for the positive measures of effort. The coefficients are all positive, and mostly increasing in magnitude, showing that reporting higher levels of these positive effort measures is correlated with higher test scores. The last two columns are for the negative effort measures, and the coefficients for these are mostly negative and increasing in magnitude.

Table 4: Correlations between discrete ordered effort measures and Spanish test scores

	Dependent variable:			
	Spanish_7			
	(Pay Attention)	(Participate)	(Miss School)	(Skip Class)
	(1)	(2)	(3)	(4)
Almost Never	0.331***	0.336***	0.005	−0.288***
	(0.055)	(0.024)	(0.008)	(0.014)
Sometimes	0.486***	0.456***	−0.355***	−0.599***
	(0.042)	(0.022)	(0.010)	(0.016)
Almost Always	0.926***	0.706***	−0.591***	−0.848***
	(0.042)	(0.022)	(0.040)	(0.031)
Always	1.044***	0.666***	−0.735***	−0.594***
	(0.042)	(0.023)	(0.043)	(0.011)
Constant	4.059***	4.425***	5.007***	5.065***
	(0.041)	(0.021)	(0.005)	(0.004)
Observations	82,244	82,108	81,653	81,178
R^2	0.047	0.024	0.021	0.054
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 9 and Figure 10 show correlations between working and reporting higher levels of the effort measures of missing school and skipping class respectively. Students who work at least one day a week report "Never missing school" less often than those who are not working. Similarly, students who are working report "Never skipping class" less often, and "Always skipping class" more often than those who are not working. Figures showing how the other measures change with work status are in Appendix A.4.

In my model, students take into consideration both the availability of schools and their

Table 5: Correlations between discrete ordered effort measures and math test scores

	Dependent variable:			
	Math_7			
	(Pay Attention)	(Participate)	(Miss School)	(Skip Class)
	(1)	(2)	(3)	(4)
Almost Never	0.253***	0.310***	−0.007	−0.257***
	(0.055)	(0.024)	(0.008)	(0.015)
Sometimes	0.306***	0.444***	−0.399***	−0.575***
	(0.042)	(0.022)	(0.010)	(0.016)
Almost Always	0.764***	0.772***	−0.532***	−0.668***
	(0.042)	(0.022)	(0.040)	(0.031)
Always	0.865***	0.744***	−0.547***	−0.544***
	(0.042)	(0.023)	(0.043)	(0.011)
Constant	4.255***	4.431***	5.041***	5.082***
	(0.042)	(0.021)	(0.005)	(0.004)
Observations	82,244	82,108	81,653	81,178
R^2	0.045	0.037	0.023	0.044
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 9: The distribution of answers for the question asking students how often they miss school, by student working status.

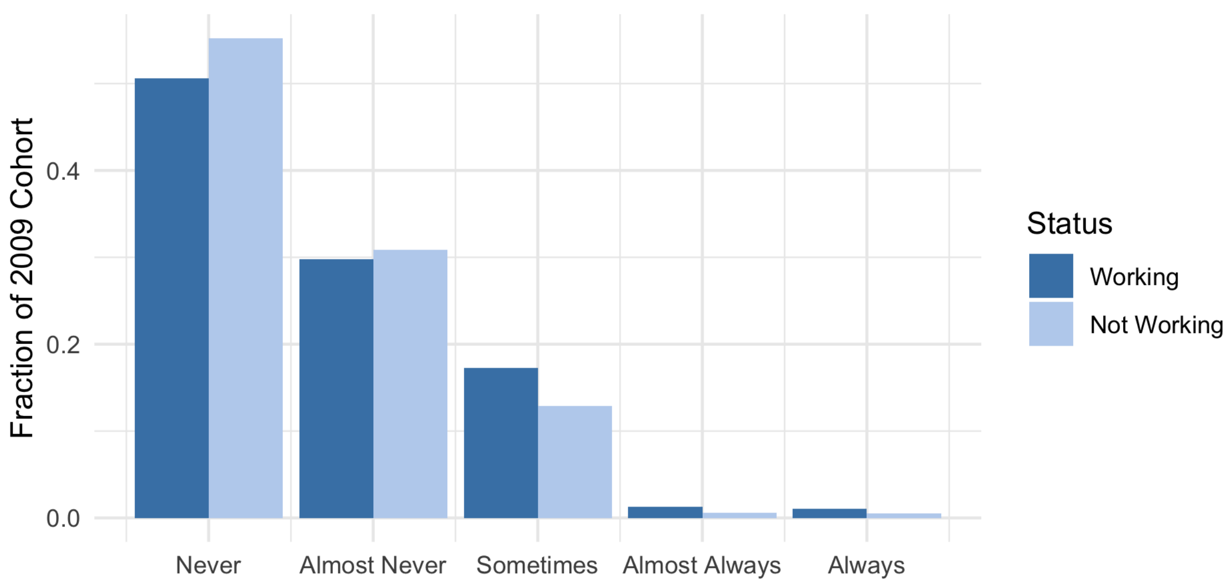
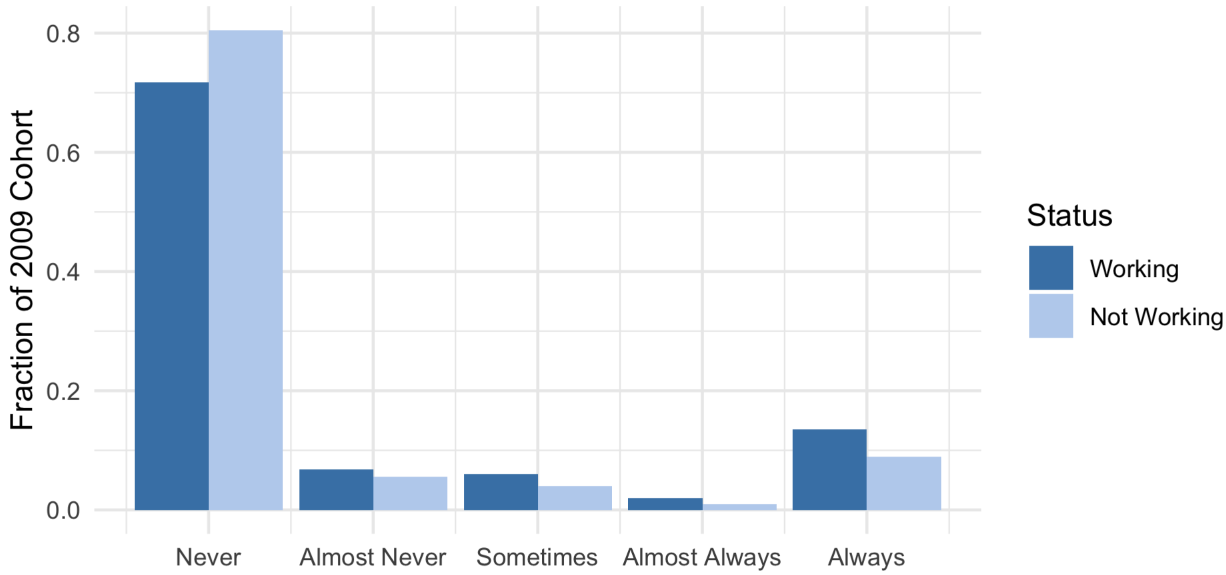
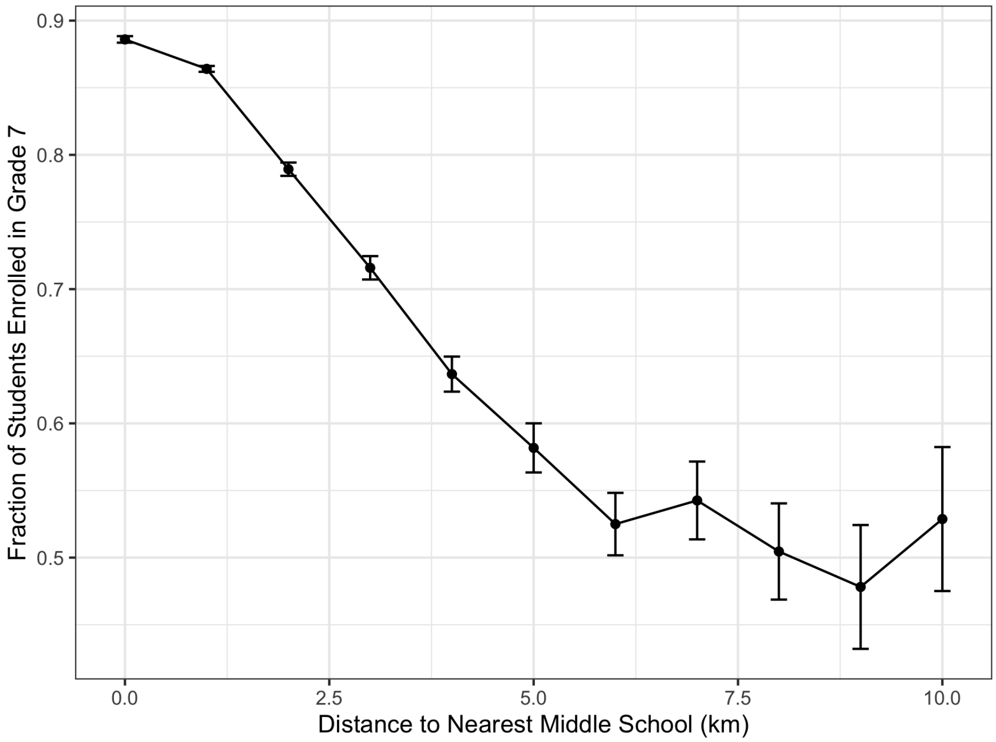


Figure 10: The distribution of answers for the question asking students how often they skip class, divided by if the students worked at least one day a week or not.



outside option of working when deciding whether to enroll in school. The farther away a middle school is from their primary school, the higher the cost of traveling there. Figure 11 shows that students who have no schools in their area are more likely to drop out than the students who have middle schools nearby.

Figure 11: The fraction of Grade 6 students in a primary school who enroll in Grade 7 as a function of how far away the nearest middle school is from their primary school.



5 Estimation

Model parameters are estimated using Maximum Likelihood. Define

$$P(j, L, A^S_{ij}, A^M_{ij}, \tilde{e}^M_{ijL} | X_i, Z_j, w_{ij}, \eta_i)$$

as the joint probability of choosing school j, labor option L, having Grade 7 test scores A^S_{ij} and A^M_{ij}, and choosing effort measures $\tilde{e}^M_{ijL}$. The probability depends on student characteristics X_i, school characteristics Z_j, imputed wages w_{ij}, and the random coefficient shock η_i. Although they are not written explicitly in the above probability, there are several other shocks in the model with defined distributions: ν_{ijL} are type I extreme value and ξ^M_{ij} and ξ^S_{ij} are jointly normal.

Using Equation 2, e^*_{ijL} can be calculated given the choice of j and L, along with the data (X_i, Z_j), the random coefficient shock (η_i) and model parameters. Define $D_{ijL} = 1$ if student

i chose school j and labor option L. The likelihood is then,

$$L = \prod_{i=1}^{N} \int \prod_{j=1}^{J_i} \prod_{L \in \{0,PT,FT\}} \left[P(j, L, A_{ij}^S, A_{ij}^M, \tilde{e}_{ijL}^M | X_i, Z_j, w_{ij}, \eta_i)\right]^{D_{ijL}} f_\eta(\eta_i) d\eta_i$$

The joint probability can be decomposed into the product of conditional probabilities. Conditioning variables in probabilities are dropped in the probability expressions if the probability does not depend on them.

$$\begin{aligned}
L &= \prod_{i=1}^{N} \int \prod_{j=1}^{J_i} \prod_{L \in \{0,PT,FT\}} \left[P(j, L, A_{ij}^S, A_{ij}^M, \tilde{e}_{ijL}^M | X_i, Z_j, w_{ij}, \eta_i)\right]^{D_{ijL}} f_\eta(\eta_i) d\eta_i \\
&= \prod_{i=1}^{N} \int \prod_{j=1}^{J_i} \prod_{L \in \{0,PT,FT\}} \left[P(A_{ij}^S, A_{ij}^M | j, L, \tilde{e}_{ijL}^M; X_i, Z_j, w_{ij}, \eta_i) P(j, L, \tilde{e}_{ijL}^M | X_i, Z_j, w_{ij}, \eta_i)\right]^{D_{ijL}} f_\eta(\eta_i) d\eta_i \\
&= \prod_{i=1}^{N} \int \prod_{j=1}^{J_i} \prod_{L \in \{0,PT,FT\}} \Bigg[P(A_{ij}^S, A_{ij}^M | j, L, ; X_i, Z_j, \eta_i) P(\tilde{e}_{ijL}^M | j, L, X_i, Z_j, \eta_i) \\
&\qquad \times P(j, L | X_i, Z_j, w_{ij}, \eta_i)\Bigg]^{D_{ijL}} f_\eta(\eta_i) d\eta_i
\end{aligned} \tag{3}$$

Consider each of the three probabilities in the likelihood. The first is the probability of observing the Grade 7 test scores in Spanish and math:

$$P(A_{ij}^S, A_{ij}^M | j, L, \tilde{e}_{ijL}^M; X_i, Z_j, \eta_i)$$

The errors for the two achievement production functions are distributed iid jointly normal. Given the choice of school and labor, the data and the model parameters, the measure of effort from the model e_{ijL}^* can be computed. Using all of these inputs, the expected test scores can be computed using Equation 1. Given the normality assumption, and the expected test scores computed from the model, the probability of observing the test scores from the data can be computed.

The second probability is the probability of observing the effort measure in the data, conditional on the optimal effort predicted from the model.

$$P(\tilde{e}_{ijL}^M | j, L, X_i, Z_j, \eta_i) = P(\tilde{e}_{ijL}^M | e_{ijL}^*)$$

In the data, there are five noisy measures of effort. One of the measures, the number of hours studied per day, is cardinal. The other four measures are ordinal variables, as they are answered on a Likert scale. To combine them into one value, I use factor analysis. This analysis is done outside of the model estimation, and uses polychoric correlations to take into account the ordinal variables. I then compute the eigenvalue decomposition of the correlation matrix, and estimate loadings for each of the five variables. The end result is a single value of effort for each student, $\tilde{e}^M_{ijL}$, which combines the information from the student's responses to the five effort questions. Estimation details and results are in Appendix A.5.

Equation 2 defines optimal effort in the model. The coefficient $\alpha_{i,9}$ in the numerator is a random coefficient with associated shock $\eta_i \sim \mathcal{N}(0, \sigma^2_\eta)$. Therefore effort draws can be thought of as coming from the distribution of the true underlying value of effort, $\mathcal{N}(e^*_{ijL}, \sigma^2_{e^*})$. This distribution is used to estimate the probability of observing the effort value obtained from factor analysis. Because of this, I do not need to simulate in order to calculate the integral defined in the likelihood.

The third and final probability is the probability of choosing school j and labor option L.

$$P(j, L|X_i, Z_j, w_{ij}, \eta_i)$$

The errors for the utility function are distributed iid type I extreme value. The probability of a school and work combination can be written as:

$$P(j, L|X_i, Z_j, w_{ij}, \eta_i) = \frac{\exp^{Utility_{ijL}}}{\sum_{k=1}^{J_i} \sum_{h \in \{0, PT, FT\}} \exp^{Utility_{ikh}}} \tag{4}$$

$Utility_{ijL}$ is a function of e^*_{ijL}, the model parameters, and the data. A scale parameter is also included in the above probability. The outside option has been normalized to the value of a wage instead of zero, and the coefficient on the monetary component is set to 1. Because of this, the scale of the distribution can be estimated.

Given a set of parameter values and the data, all three of these probabilities can be calculated for each student, and the product of them is defined as the individual likelihood. The likelihood defined in Equation 3 can then be calculated, and maximized to find the estimated parameters.

5.1 Identification

There are 51 parameters to estimate in the model in total. The list of parameters is given by

- Utility function: $\{\alpha_k\}_{k=1}^{11}$, $\{\beta_k\}_{k=1}^{2}$, $\{\lambda_k\}_{k=1}^{3}$, σ_U
- Achievement production functions: $\{\delta_k^M\}_{k=1}^{15}$, $\{\delta_k^S\}_{k=1}^{15}$, σ_M, σ_S, σ_{MS}
- Effort: σ_E

There are 33 parameters associated with achievement. They are estimated with two value added equations. Each student who attended Grade 7 has a test score in both Math and Spanish. Each student also has lagged test scores in both subjects, as well as data on the 12 other covariates. There is variation in covariates across schools, and across students within a school.

There are 16 parameters in the utility function. Two of the parameters are associated with distance. They are identified by geographic variation in distances in different children's choice sets. Each primary school has different schools in its choice set, and every option is associated with a distance (among other characteristics). Schools that are far away from a specific primary school may be of a good quality, but are chosen by a small fraction of the students (or not at all), which identifies how costly students find traveling to school.

Three parameters in the utility function represent school type (General, Technical, Telesecondary). There are too many schools in the data to have intercepts for each of them. Instead of having a common intercept in the utility for attending each school, I assume that the intercept varies by school type. These coefficients are identified by variation within choice sets as well. Students may chose a certain type of school over another even though it is farther away or offers a worse expected test score, showing a preference for this type of school over the other.

Two parameters in the utility function capture how much students value expected test scores. Two factors come into play here. The first is that students with higher test scores may get more utility from going to school compared to dropping out. The second is that achievement is affected by school inputs, so some schools in the choice set may have higher expected test scores which could make students more likely to attend. Either of these things being present in the data would identify the coefficients on test scores.

There are six parameters associated with the marginal cost of effort in the utility function. The parameters involved in demographics (parental education, female, lagged test scores) are identified by the difference in mean effort choices from students with these different demographics.

6 Results

Estimates for the utility parameters are shown in Table 6 and for the test score production functions are shown in Table 7. All parameters in the utility function have the units of 100s of pesos per month. The key parameter estimates and patterns are discussed below. Traveling distance to a middle school is estimated to be costly. The coefficient on distance squared is positive, showing that as the school gets farther away, the marginal cost of another kilometer starts to decrease. Both estimates are significant, even with a small estimation sample compared to the full data sample.

Coefficients	Estimates	Std.Error
Distance	-46.49	3.22
Distance squared	1.90	0.13
School	36.97	17.98
School x Parent Educ	6.42	3.90
Technical School	11.26	2.43
Telesecondary School	-37.24	3.99
Expected Score	28.04	3.09
Expected x Parent Educ	0.39	2.77
Expected x CCT	0.42	0.13
Working Part Time	-56.62	3.71
Linear Effort	-33.81	9.62
Linear Effort - Lagged Score	1.86	0.29
Linear Effort - Female	0.94	0.30
Linear Effort - Parent Educ	-0.53	6.15
Linear Effort - Work	-0.43	0.19
Quadratic Effort	-5.23	0.86

Table 6: Coefficient estimates for parameters in the utility function. Estimates come from a sample of 10,000 students.

The estimate for attending school in the utility is large and positive, and does not seem to depend on if parents have middle-school education. Technical schools are estimated to be slightly more valuable than general schools, but the difference is not significant. Telesecondaries are estimated to be perceived significantly worse than the other two school types.

The average expected test score has a positive coefficient in the utility function, with little change depending on parental education and conditional cash transfer status. The standard deviation of the test scores is approximately 1, meaning that students and their families place approximately the same value on a school being a kilometer closer as the school improving math test scores by one and a half standard deviations.

There is a large distaste for working part time. Further, working part time is estimated to make the marginal cost of effort more negative, so more costly. The coefficient on effort squared must be negative to guarantee a solution to the optimal effort problem in the model, and it is in fact a large negative number. The marginal cost of effort is estimated to decrease, so effort is less costly, for female students and students with higher lagged test scores.

Coefficients	Math_Estimates	Math_Std.Error	Spanish_Estimates	Spanish_Std.Error
Intercept	-0.712	0.455	-0.511	0.493
Lagged Math	0.262	0.032	-0.06	0.03
Lagged Spanish	0.044	0.031	0.305	0.032
Female	-0.173	0.034	0.133	0.034
Age	-0.084	0.014	-0.072	0.014
Technical School	0.069	0.122	0.124	0.13
Telesecondary School	0.363	0.192	-0.815	0.21
Principal Education	-0.003	0.007	-0.026	0.007
Principal Experience	0.001	0.002	0.002	0.002
School has Internet	0.175	0.019	0.151	0.019
School has Materials	-0.034	0.015	-0.005	0.015
Teachers are Bad	-0.037	0.009	-0.037	0.009
Effort	1.12	0.156	1.103	0.156
Effort X Technical	-0.003	0.024	-0.019	0.025
Effort X Tele	-0.065	0.035	0.151	0.037
Residual Std Error	0.537		0.545	
Residual Covariance	0.253			

Table 7: Coefficient estimates for parameters in the achievement functions. Estimates come from a sample of 10,000 students.

The coefficient estimates in the achievement production function, shown in Table 7, are fairly intuitive. Lagged test scores are significant, with lagged math scores contributing to math predictions, and lagged Spanish scores contributing to Spanish predictions. Females have negative coefficients in math. Students with higher ages are estimated to do worse in both math and Spanish. The value added of a Technical schools is estimated to be greater than a General school in both math and Spanish whereas Telesecondaries are estimated to be worse. The school characteristics coefficients are mainly small in magnitude and insignificant. The one exception is if the school has internet, which has a positive and significant coefficient. Finally, effort has a large positive coefficient for both math and Spanish. Effort is estimated to be more productive in Telesecondary schools compared to General schools, and less productive in Technical schools, however the change in productivity is small in magnitude.

6.1 Model Fit

The following figures show the fit of the model with respect to the true data. There are three main outcomes to fit: school choice, achievement, and effort. Table 8 shows the model fit for the means and standard deviations of these outcomes. The simulation means are overall quite close to the means in the data.

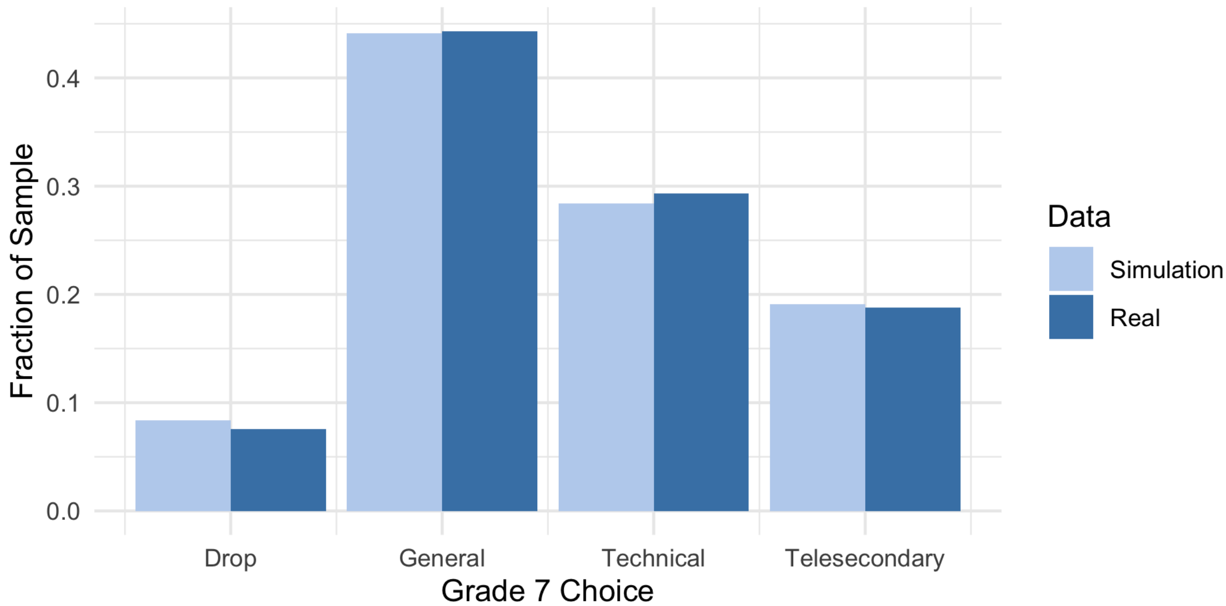
Outcome Variable	True Mean	Simulated Mean	True St.Dev.	Simulated St.Dev.
Math	5.03	5.28	0.98	1.59
Spanish	5.00	5.26	0.98	1.61
Effort	4.66	4.67	1.28	1.26
Fraction Drop	0.08	0.08		
Fraction General	0.44	0.44		
Fraction Technical	0.29	0.28		
Fraction Telesecondary	0.19	0.19		
Fraction Work PT	0.25	0.22		

Table 8: Model fit for relevant means and standard deviations

Figure 12 shows the fraction of students that choose each of the three school type options or to drop out in both the simulation and the data. The pattern in the data is represented in the simulated data, in that General schools are most popular, followed by Technical, Telesecondaries and then Dropping out. However the values are slightly off, with somewhat more students choosing to drop out in the model than in the data.

Considering only the students who choose to drop out, Figure 13 investigates the relationship between dropping out and distance to the nearest school. Students are divided into quintiles by the distance to their nearest middle school. The mean dropout rate for each quintile is then calculated in the data and the model. The overall pattern matches, but it is apparent that the model is overestimating dropout rates for students who have a middle

Figure 12: The fraction of students choosing each of the three school types or dropping out in Grade 7 both in the data and in the estimated model.



school in the same location as their primary school, or who have a middle school very far away. The students in between are matched very closely.

Finally, Figure 14 investigates the relationship between students' effort values and their lagged test scores. In both the data and the model, students with lower lagged test scores exert less study efforts than students with higher lagged test scores. The model captures the relationship in the data very well.

Figure 13: The fraction of students who dropout broken down by how far away the nearest school is from their primary school.

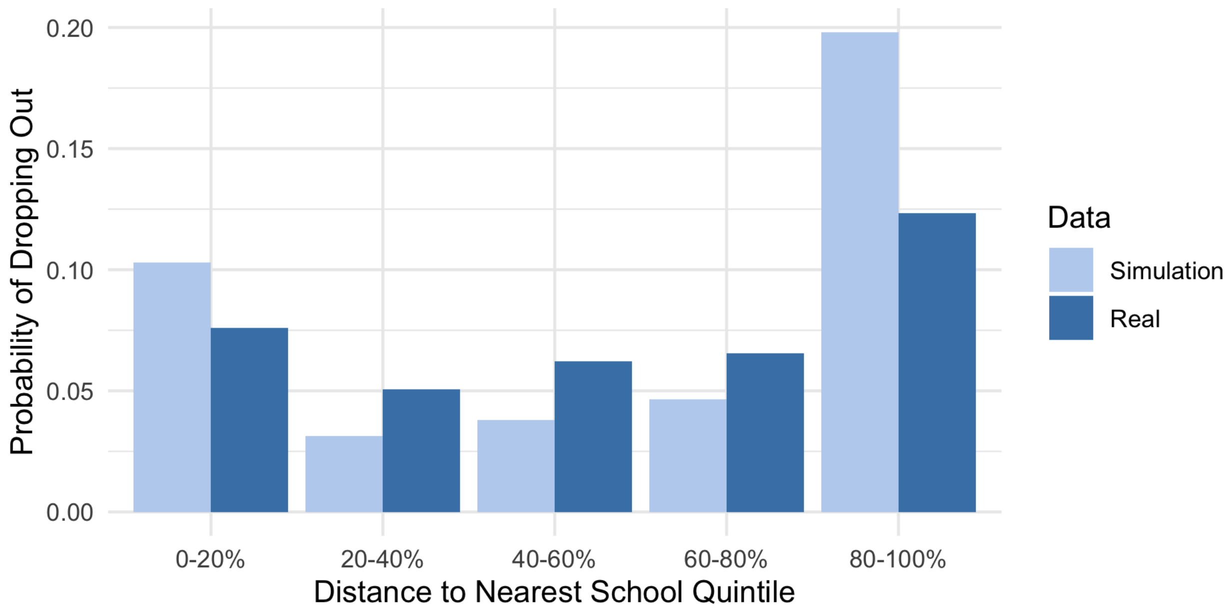
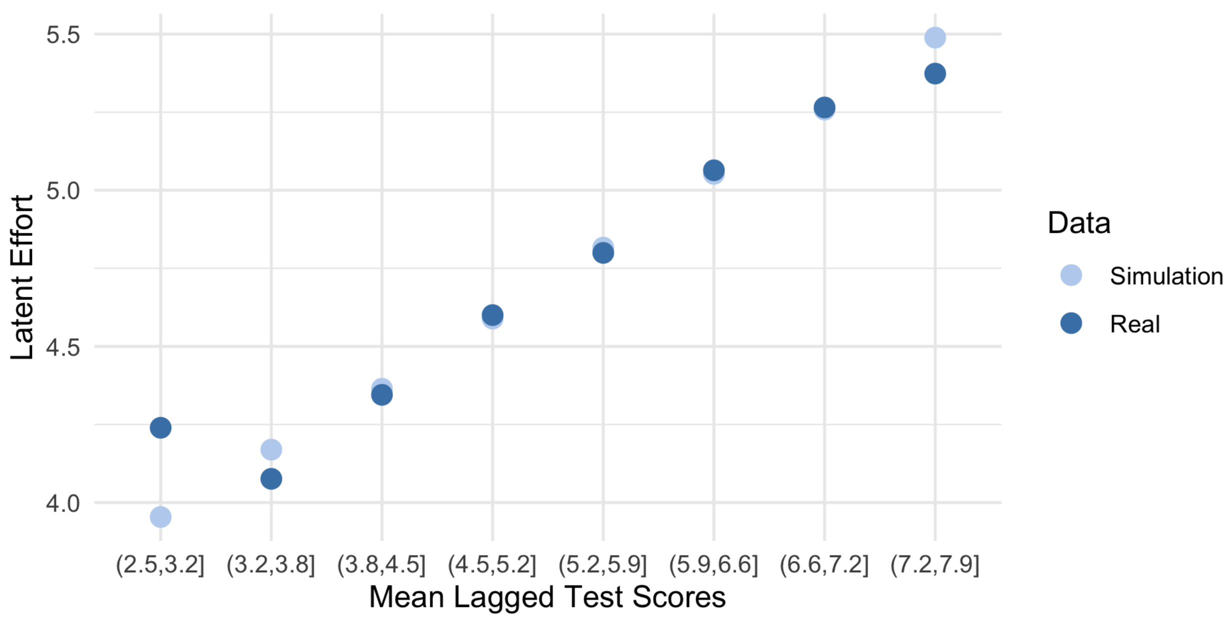


Figure 14: The effort variable from the data and the effort variable generated by the model are plotted as a function of average lagged test scores.



7 Evaluation of Child Labor Policies

With my estimated model, I am able to evaluate many relevant policies involving child labor laws, conditional cash transfers and school availability. The focus for this paper is to consider the impact of enforcing child labor laws on both dropout and achievement. Working while enrolled in school is detrimental to achievement, however I find that for many students they require the income to stay in school, and if they are not allowed to work while in school they prefer to drop out and work full time. There are two ways to counter this problem. The first is to fully prohibit child labor, both while enrolled in school and if the child has dropped out. This makes the outside option less appealing and more children will stay in school. The second is to offer conditional cash transfers as an incentive for students to enroll. The cash transfers may be the more feasible policy, however program targeting can still pose a challenge and affect the results, as does the benefit amount offered.

Using the parameter estimates, I draw shocks and simulate choices under the baseline model. Then, to do the counterfactual exercises, I change either some parameters or the choice sets that the students face, and simulate again under the modified environment using the same shocks. The results from the two simulations are compared to evaluate the policy. Of interest are the change in enrollment rates, the change in achievement, which types of schools have the largest change in enrollment, the amount of money gained/lost by families, among other outcomes.

The first counterfactual involves removing the part-time labor option. The students who chose not to work originally are not affected by this policy, and neither are students who chose to drop out. However, the children who were working while enrolled in school must decide if they wish to continue studying without the income they received, or drop out and work full time. This counterfactual could represent a policy such as teachers being able to better monitor their students, or if there was an after-school program implemented so that children studied or played sports at school during hours when they may normally work. The results of this policy are shown in Table 9. In the estimated model, 22% of students enrolled in school choose to work at the same time. When working part time is not an option 7% of these students decide to dropout, increasing the over all dropout rate by 20%, from 8% to 10%. This is a drastic increase, and represents a large number of students when considering the entire student population of Mexico. There does not appear to be much change in effort, Spanish and Math scores overall, so I will investigate these further.

Outcome Variable	Estimated Model	Counterfactual	Percent Change
Fraction Work PT	0.22	0.00	-100.00
Fraction Drop	0.08	0.10	19.28
Mean Effort	4.67	4.68	0.26
Mean Spanish	5.26	5.28	0.40
Mean Math	5.28	5.30	0.38

Table 9: Changes in outcomes from Counterfactual 1.

Table 10 computes statistics for the group of students who would like to work, but when they are prohibited from working, stay enrolled in school. From this table, it is clear that they are increasing their efforts, which in turn increases their math and Spanish scores. Effort increases by an average of 3.3 percent of a standard deviation, which results in a 2.9 percent of a standard deviation increase in math scores and a 3 percent of a standard deviation increase in Spanish scores.

Outcome Variable	Estimated Model	Counterfactual	Change in SD (%)
Effort	4.65	4.69	3.27
Math	5.29	5.33	2.92
Spanish	5.27	5.32	2.96

Table 10: Changes in outcomes from Counterfactual 1 for students who would like to work part time, but stayed enrolled when they could not.

The final analysis analyzes the characteristics of the students who are most likely to drop out because of this policy. Table 11 shows the mean value of background characteristics for the students who would like to work while in school, separated by if they stay in school or not after the policy. The students who drop out have lower lagged test scores and have a much higher rate of being a conditional cash transfer beneficiary. The gender breakdown is very similar and the students that dropout are only slightly older than those that stay enrolled. The last three rows show that the students who stay enrolled are much more likely to have one of their parents have at least a middle school education, and to be in the top half of the income distribution. Overall, this table shows that the students who are dropping out are the students who are struggling academically and come from disadvantaged backgrounds.

Increasing the national dropout rate by such a large amount is not an ideal result of a policy that prohibits working while in school. A possible way to prevent this, would be to consider prohibiting all labor. This would reduce the value of dropping out, as the students who dropped out would not receive wages. Results from this counterfactual, along with the

Background Variables	Stayed In School	Dropped Out
Lagged Math	5.17	4.55
Lagged Spanish	5.16	4.51
Female	0.51	0.46
CCT	0.22	0.47
Age	12.00	12.26
Mom has Middle School	0.56	0.32
Dad has Middle School	0.58	0.33
Family Income above Mean	0.43	0.25

Table 11: Difference in background variables between the students who dropped out when prohibited from working part time, and the students who stayed enrolled.

results from the estimated model and the first counterfactual are shown in Table 12. The numbers show that prohibiting all child labor would have better impacts than only prohibiting labor while in school, in that the dropout rate is similar to the original estimation. Effort and test scores are also slightly higher than in the baseline model, since students are dedicating all of their time to their studies.

Outcome Variable	Estimated Model	Counterfactual 1	Counterfactual 2
Fraction Work PT	0.22	0.00	0.00
Fraction Drop	0.08	0.10	0.08
Mean Effort	4.67	4.68	4.68
Mean Spanish	5.26	5.28	5.26
Mean Math	5.28	5.30	5.29

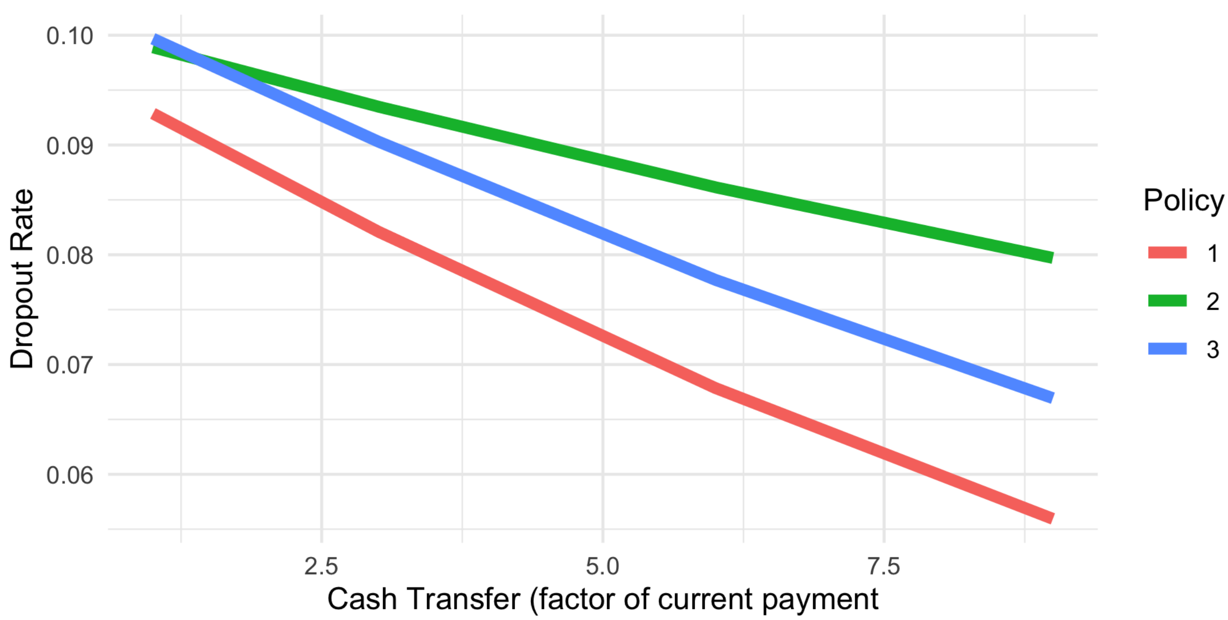
Table 12: Outcomes from the estimated model, Counterfactual 1 and Counterfactual 2.

Although prohibiting all child labor may have positive educational outcomes, as a policy it would be difficult to enforce. Therefore I look to an alternative policy to encourage enrollment if working part time is prohibited. Luckily in Mexico there is a well established policy, the conditional cash transfer, that could be modified. In my third counterfactual, I consider changing the values of the conditional cash transfer and expanding eligibility for the program.

Figure 15 shows the reduction in dropout rates for three different conditional cash transfer policies. The x-axis shows an increase in the amount of the transfer, ranging from the current Prospera transfer amount, to 9 times the current amount. The three policies change who is offered the conditional cash transfer. Policy 1 is a hypothetical policy that is not operational, but shows the best that could be achieved with a cash transfer of the given magnitude. In this policy, any student who would drop out in counterfactual 1 is offered the transfer. In reality, it would be impossible to target the policy this way. Policy 2 considers increasing the transfers to the current beneficiaries, which would be very simple to implement. Policy 3 extends the transfer beneficiaries to those who currently received Prospera, and those who have an income below the median.

The results show that increasing the conditional cash transfer payment is a very effective way of decreasing the dropout rate. For payment amount similar to the current value, expanding the conditional cash transfer to other low-income families does not have a significant effect. However, if the cash transfer is increased, then extending the transfer to these families does drastically help reduce the dropout rate.

Figure 15: The fraction of students who dropout when considering three different conditional cash transfer policies. Policy 1 offers the transfer to any student who wants to drop out. Policy 2 offers the transfer to current beneficiaries. Policy 3 offers the transfer to current beneficiaries and students whose family earns below the median income.



To summarize the results, I find that prohibiting students from working while in school increases achievement by approximately 3% of a standard deviation, however it also causes a substantial increase in the dropout rate. If it were possible to ban all child labor, the dropout rate would remain close to the baseline and achievement would increase. However, this would be difficult to enforce, and I find that similar dropout rates can be achieved when working part time is banned and the cash transfer is either increased, or the beneficiaries expanded.

8 Conclusion

Increasing human capital is thought to be one of the best ways for developing countries to achieve growth and to increase equity. Ensuring that all children attend school to a certain age and receive a high quality education is a priority. Unfortunately, in many developing countries, child labor is prevalent and it makes providing an education to all students more challenging. Although there is an extensive literature on school choice, it is necessary to extend the currently available frameworks to consider the problem of child labor and how it interacts with school choices. In my model, I include both schooling and labor choices and I provide a mechanism through which labor affects educational achievement, which is the study effort that children dedicate to their education.

Specifically, I develop and estimate a random utility model over discrete school-work alternatives, where study effort is determined as the outcome of an optimization problem under each of these alternatives. Students who do not enroll in school are assumed to work full-time, and receive the associated wage. Students who enroll in school may choose to work part-time, for which they receive the benefit of a part-time wage, but incur the cost of increased marginal cost of effort. The results show that effort is an important input to achievement, which is estimated with a value added equation. Students who work, and as a result choose to put in less, end up with lower achievement than they would if they had not chosen to work.

To estimate my model, I combine several data sources: administrative data on nationwide standardized tests in math and Spanish, survey data from students, parents and principals, geocode data on school locations, and Mexican census data on local labor market wages and hours worked. The majority of the model parameters are precisely estimated.

By removing the part-time labor option from student's choice sets, I evaluate the impact of working while in school. I find that for the majority of students, not being able to work improves their test scores. However, almost 10% of students who would prefer to work drop out of school to work full time when the part time option is no longer available. This increases the dropout rate by approximately 20%. I analyze two policies that could be used in conjunction with prohibiting labor for enrolled students. The first is to ban all child labor. This removes the incentive to drop out of school and work full time and reduces the dropout rate, however it would be a challenging policy to implement. The second policy is to increase the conditional cash transfer, in both the payment amount and the pool of beneficiaries. I find that depending on the transfer amount, these policies show considerable potential.

With the model that I have developed and estimated, it is possible to analyze many other educational policies. I incorporate school choice and locally available schools, so one possible direction is to consider questions of school access and quality. Especially in rural areas, it is of interest to understand how the conditional cash transfer interacts with another important education policy in Mexico, the distance education schools (Telesecondaries). In ongoing work, I am considering a range of such policies.

References

Acevedo, I., Ortega, A., Székely, M., 2019. Rendimiento Escolar y Transiciones Laborales con Transferencias Condicionadas en México. Working Paper .

Attanasio, O.P., Meghir, C., Santiago, A., 2011. Education choices in Mexico: using a structural model and a randomized experiment to evaluate Progresa. The Review of Economic Studies 79, 37–66.

Behrman, J.R., Parker, S.W., Todd, P., 2019. Impacts of Prospera on Enrollment, School Trajectories and Learning. Working Paper .

Behrman, J.R., Sengupta, P., Todd, P., 2005. Progressing through PROGRESA: An impact assessment of a school subsidy experiment in rural Mexico. Economic development and cultural change 54, 237–275.

Bourguignon, F., Ferreira, F.H., Leite, P.G., 2003. Conditional cash transfers, schooling, and child labor: Micro-simulating brazil's bolsa escola program. The World Bank Economic Review 17, 229–254.

Dubin, J.A., McFadden, D.L., 1984. An econometric analysis of residential electric appliance holdings and consumption. Econometrica: Journal of the Econometric Society , 345–362.

Dubois, P., De Janvry, A., Sadoulet, E., 2012. Effects on school enrollment and performance of a conditional cash transfer program in Mexico. Journal of Labor Economics 30, 555–589.

Epple, D., Jha, A., Sieg, H., 2018. The superintendent's dilemma: Managing school district capacity as parents vote with their feet. Quantitative Economics 9, 483–520.

Gottfried, M.A., 2010. Evaluating the relationship between student attendance and achievement in urban elementary and middle schools: An instrumental variables approach. American Educational Research Journal 47, 434–465.

Heckman, J.J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models, in: Annals of economic and social measurement, volume 5, number 4. NBER, pp. 475–492.

Keane, M.P., Krutikova, S., Neal, T., 2018. The impact of child work on cognitive development: results from four Low to Middle Income countries. Technical Report. IFS Working Papers.

Leite, P., Narayan, A., Skoufias, E., 2015. How do ex ante simulations compare with ex post evaluations? evidence from the impact of conditional cash transfer programs. Journal of Poverty Alleviation & International Development 6.

Neilson, C., 2013. Targeted vouchers, competition among schools, and the academic achievement of poor students. Job Market Paper , 1–62.

Parker, S.W., Todd, P.E., 2017. Conditional cash transfers: The case of Progresa/Oportunidades. Journal of Economic Literature 55, 866–915.

Schultz, T.P., 2004. School subsidies for the poor: evaluating the Mexican Progresa poverty program. Journal of development Economics 74, 199–250.

Stinebrickner, R., Stinebrickner, T.R., 2008. The causal effect of studying on academic performance. The BE Journal of Economic Analysis & Policy 8.

Todd, P., Wolpin, K., 2006. Assessing the impact of a school subsidy program in mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. The American economic review 96 5, 1384–1417.

Todd, P., Wolpin, K.I., 2018. Accounting for mathematics performance of high school students in mexico: Estimating a coordination game in the classroom. Journal of Political Economy 126, 2608–2650.

Yap, Y.T., Sedlacek, G., Orazem, P.F., 2009. Limiting child labor through behavior-based income transfers: An experimental evaluation of the peti program in rural brazil.

A Appendix

A.1 Wage Regressions

The data used to estimate the wage regressions comes from the Mexico 2010 Census, and can be accessed through the IPUMS site: https://international.ipums.org/international-action/variables/search. The variables that are downloaded are:

- Age of subject (MX2010A AGE)
- Whether or not the subject currently attends school (MX2010A_SCHOOL)
- Income of individual for the last month (MX2010A_INCOME)
- Household's income from work (MX2010A_INCHOME)
- Number of hours worked by individual in the last week (MX2010A_HRSWORK)
- Educational attainment level of individual in number of years (MX2010A_EDATTAIN)
- Educational attainment level of individual in number of years MX2010A_EDATTAIN_MOM)
- Educational attainment level of individual in number of years (MX2010A_EDATTAIN_POP)
- Gender (MX2010A_SEX)
- Employment status (MX2010A_EMPSTAT)
- Position at work (MX2010A_CLASSWK)
- State code (GEO1_MX2010)
- Municipality code (GEO2_MX2010)
- Urban-rural status (URBAN)

In order to compute the regressions, we recode several new variables from the ones listed above:

- **INCOME_PER_HOUR** (Created by dividing income last month by 4 times the number of hours worked last week)
- **familyworker** (Dummy for whether the individual is an unpaid family worker)

- **mom_edattain_missing** (Created from MX2010A_EDATTAIN_MOM variable; 1 = mom's educational attainment is missing, 0 = mom's educational attainment is not missing)
- **dad_edattain_missing** (Created from MX2010A_EDATTAIN_POP variable; 1 = dad's educational attainment is missing, 0 = dad's educational attainment is not missing)
- **north_dummy** (Dummy for whether municipality is in the North or South region of Mexico ; 1 = North, 0 = South)

The next steps are to clean and filter the data:

1. Exclude individuals with an undefined age (include only MX2010A_AGE != 999)
2. Exclude individuals with undefined school attendance status (include only MX2010A_SCHOOL == 1 |MX2010A_SCHOOL == 2)
3. Assign 0 to missing or unknown values for monthly personal income (MX2010A_INCOME), monthly family income (MX2010A_INCHOME), and hours worked in the last week (MX2010A_HRSWORK)
4. Create Income Per Hour variable by dividing MX2010A_INCOME (monthly income) by 4 times MX2010A_HRSWORK (hours worked in the last week) and assign infinite and undefined values to 0
5. Reassign educational attainment variables (MX2010A_EDATTAIN, MX2010_EDATTAIN_MOM, MX2010A_EDATTAIN_POP) values with continuous values
6. Create mom_edattain_missing and dad_edattain_missing variables by assigning a 1 for these variables if the MX2010A_EDATTAIN_MOM and MX2010A_EDATTAIN_POP are missing or unknown, respectively, and a 0 if not
7. Include only individuals with educational attainment levels equal to or below 13 (MX2010A_EDATTA <= 13) in order to exclude students who have finished high school
8. Create a dummy variable (mun_dummy) that indicates whether (1) or not (0) the municipality the individual is in also contains a city with at population of at least 100,000 (merged with CityCoordinates_withMunicipalities file)

9. Create a north/south dummy (north_dummy) that indicates whether the municipality is in the northern (1) or southern (0) region of Mexico

10. Filter by age to only include inviduals between the age of 12 and 20 inclusive (MX2010A_AGE <= 20 & MX2010A_AGE >= 12)

11. Create cutoffs for INCOME_PER_HOUR, MX2010A_HRSWORK, MX2010A_INCHOME and exclude entries for each variable with values above the 99th quantile

12. Create a yeswork variable where yeswork = 1 if one of the following criteria are met:

    - MX2010A_EMPSTAT == 10
    - MX2010A_HRSWORK != 0
    - MX2010A_INCOME != 0
    - MX2010A_CLASSWK == 1
    - MX2010A_CLASSWK == 2
    - MX2010A_CLASSWK == 3
    - MX2010A_CLASSWK == 4
    - MX2010A_CLASSWK == 5
    - MX2010A_CLASSWK == 6

    and MX2010A_HRSWORK >= 5 and INCOME_PER_HOUR >0

13. Create family net income (netincome) variable by subtracting individual's income from their entire family's income (which includes the individual's income): MX2010A_INCHOME - MX2010A_INCOME

14. Then create dummies for family income where

    - family_income1 includes netincome <1500
    - family_income2 includes 1500 <= netincome <3000
    - family_income3 includes 3000 <= netincome <7500
    - family_income4 includes 7500 <= netincome <15000
    - family_income5 includes 15000 <= netincome <30000
    - family_income6 includes netincome >= 30000

15. Create separate nonzero_data data set by filtering yeswork ==1

16. Separate into two data sets based on gender

Wage regressions are estimated on the two data sets separately, using a Heckman selection model. The first step is to run a probit model on the probability of working. The full dataset is used to estimate this probit. The wage regressors include: age, school attendance, educational attainment, parental educational attainment, parents are missing, urban-rural dummies, north-south dummies, and municipality dummies. In addition, the following variables are assumed to influence selection into working, but not the wage offers, and are included as exclusion restrictions: family income, home electricity, home piped water, home internet and home computer.

Using the results from the probit, it is possible to create control functions for each student. These are included as a regressor in the next step of the estimation process, which is a fixed effect linear regression model with hourly wages as the independent variable, and the regressors listed in the probit (without the exclusion restriction variables). The fixed effects in the model are the municipality fixed effects, which allow for great geographic heterogeneity. This regression is estimated using the subset of students who report working and earning a positive wage.

Results for the wage regressions are shown in Table.

To impute wages for all students in the estimation sample, I use the coefficients from the wage regressions. With this, I get an estimate of the hourly wage for each student. A separate regression was estimated for number of hours worked per week. Instead of imputing the hours worked per week for each student (which I will incorporate in the future), I use the mean hours worked by gender and school attendance status. These hours are shown in Table 13. For each student, I multiply their hourly wage by the number of hours worked and by four in order to estimate their monthly wage offer. Figures 16 and 17 shows the imputed wages for the sample.

	Male	Female
In School	16.52	15.79
Out of School	37.15	35.60

Table 13: Estimated hours worked per week for 12 year old children using census data.

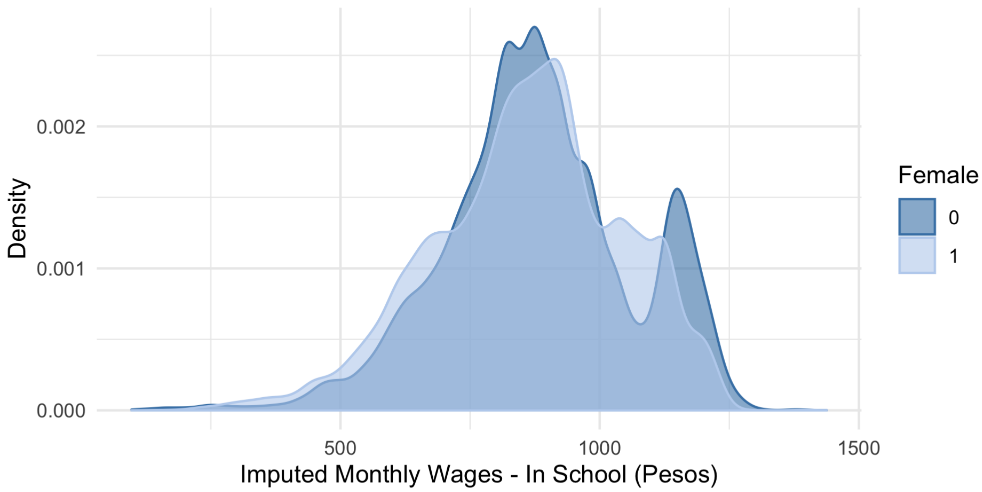


Figure 16: The distribution wage offers for male and female students who are enrolled in school.

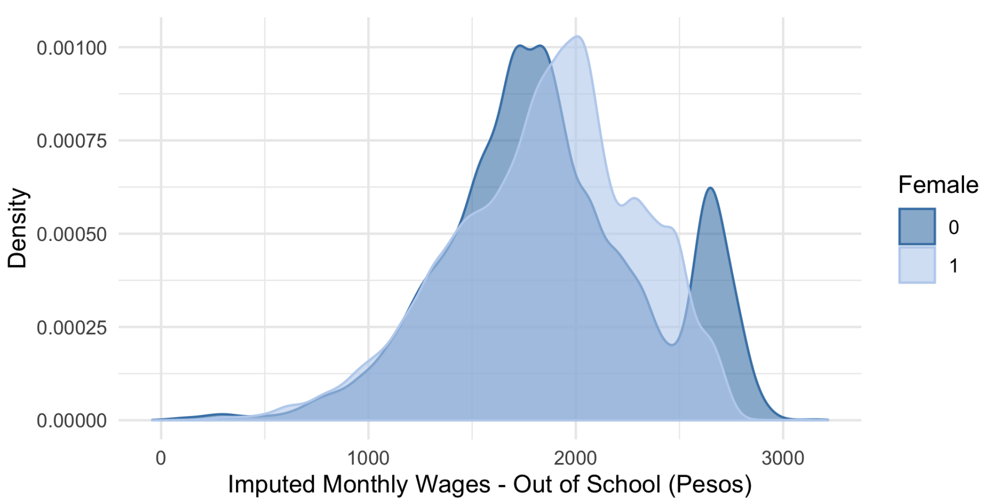


Figure 17: The distribution wage offers for male and female students who are not enrolled in school.

A.2 Creating Estimation Sample

Dropped students wrote survey in Grade 6.
Enrolled students wrote survey in Grade 7.
Background information assumed to stay constant.
Need effort from the Grade 7 survey.
Need to make sure that the ratio of dropped to enrolled is the same as in the raw data.

A.3 Histograms of Raw Test Scores

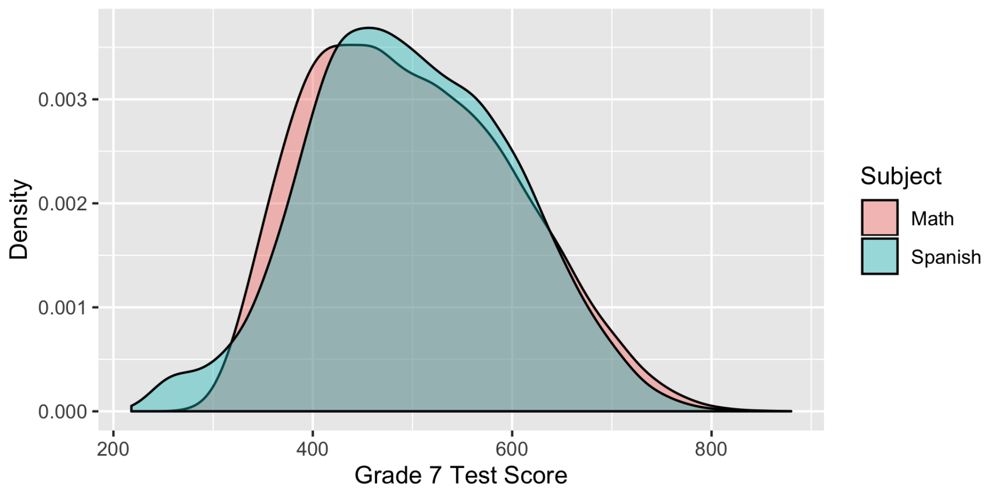


Figure 18: The distribution of Grade 7 test scores in 2009 in the estimation sample.

A.4 The Effect of Working on Effort Question Responses

A.5 Factor Analysis for Effort Questions

I use factor analysis to estimate the latent effort variable. I am assuming that there is a true unobserved latent effort variable, and that the five questions that I observe are all affected by the latent variable. Formalizing this, I assume that the unobserved latent effort variable $\hat{e}_i^M$ is connected to the five measures in the data $(e_{i1}^M, ..., e_{i5}^M)$ in the following way,

$$\begin{aligned} e_{i1}^M &= \gamma_1 \hat{e}_i^M + u_{i1} \\ &\vdots \\ e_{i5}^M &= \gamma_5 \hat{e}_i^M + u_{i5} \end{aligned}$$

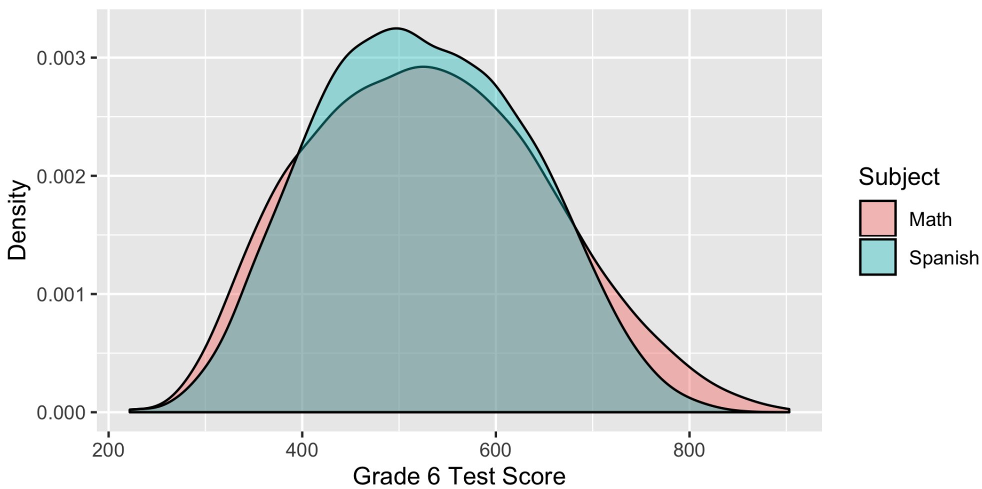


Figure 19: The distribution of Grade 6 test scores in 2008 in the estimation sample.

Figure 20: The distribution of answers for the question asking students how often they pay attention in class, divided by if the students worked at least one day a week or not.

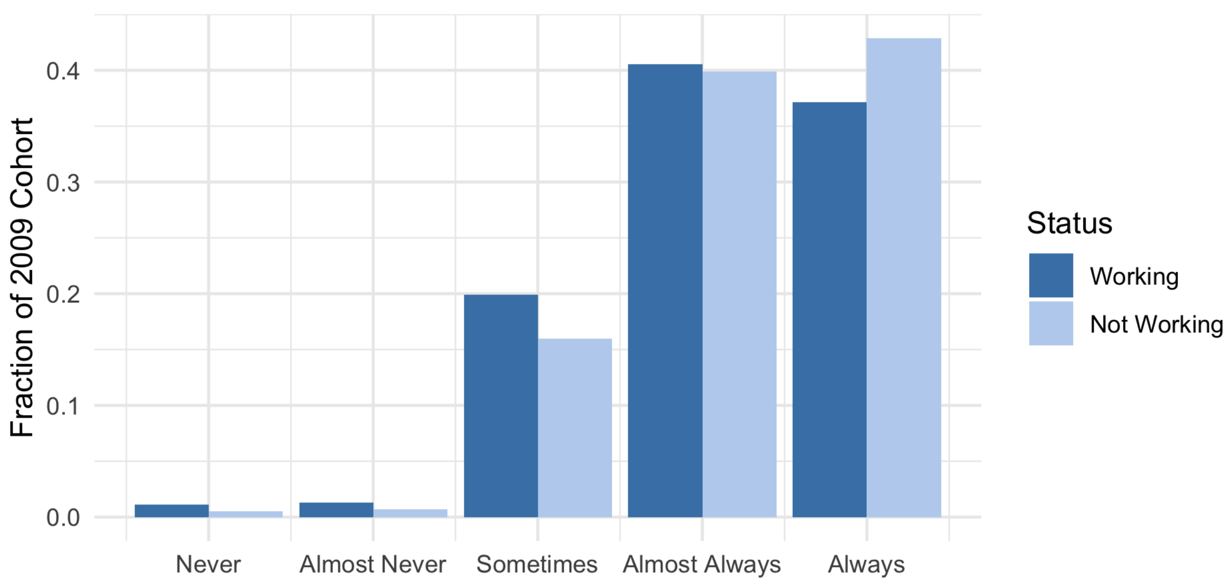


Figure 21: The distribution of answers for the question asking students how often they participate in class, divided by if the students worked at least one day a week or not.

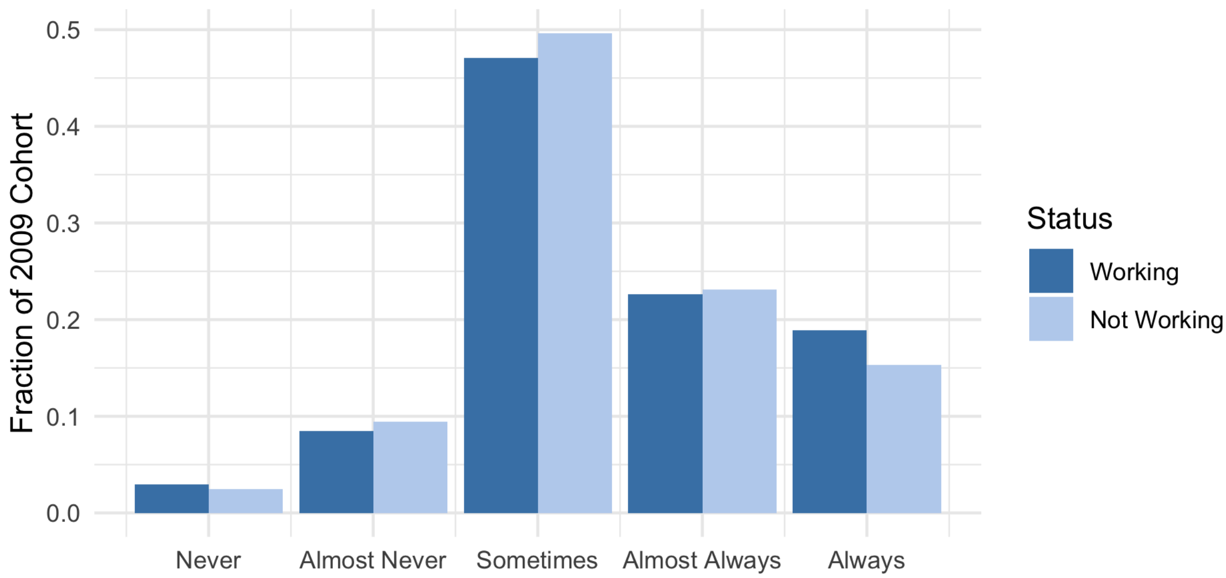


Table 14: Correlation between working and the number of hours studied per day

	Dependent variable:
	StudyHours
Working	-0.034^{***}
	(0.008)
Constant	1.768^{***}
	(0.004)
Observations	82,580
R^2	0.0002
Note:	*p<0.1; **p<0.05; ***p<0.01

First, I compute the correlation matrix of the five measures in the data. Because four of the measures are ordinal variables, I compute a polychoric correlation matrix. This follows the practice in the literature, and the main assumption is that the ordinal variables have an underlying joint continuous distribution. The polychoric correlation matrix for my five measures of effort is calculated to be:

	Pay Attention	Participate	Miss School	Skip Class	Study Hours
Pay Attention	1.00				
Participate	0.43	1.00			
Miss School	-0.23	-0.13	1.00		
Skip Class	-0.24	-0.14	0.25	1.00	
Study Hours	0.28	0.20	-0.11	-0.08	1.00

The signs of the correlations are as would be expected, with paying attention in class, participating in class and the number of hours studied per day all positively correlated with each other, and negatively correlated with missing school and skipping class.

To compute the factor loadings and get an estimate for the latent effort variable I use the Principal Axis method. This is an iterative procedure, and iterates until the communalities of each of the measures do not vary by iteration. Communalities are defined as the component of the variance of each of the measures that are shared, and therefore can be attributed to the latent factor. The initial guess of the communality of a given variable comes from the R^2 of the regression using that variable as the independent variable, and the other four measures as the dependent variables. These initial guesses replace the diagonal elements of the correlation matrix. Then, an eigendecomposition is done of this updated correlation matrix. Using the eigenvalues and eigenvectors, new communalities can be computed. This is repeated, until the communalities stabilize. After convergence, the loadings are extracting using the eigenvalues and eigenvectors of the final matrix.

The loadings for each of the effort variables are,

Variables	Loadings
PayAttention	0.77
Participate	0.53
MissSchool	-0.34
SkipClass	-0.34
StudyHours	0.35

To get an estimate of the latent effort variable $\hat{e}_i^M$ for each student i, I multiply their effort measures by the associated loading factor.

$$\hat{e}_i^M = l_1 * e_{i1}^M + ... + l_5 * e_{i5}^M$$

The result is a continuous effort variable for each student, that has greater variance than any of the individual measures used to compute it. Figure 22 shows a histogram of the final effort measures.

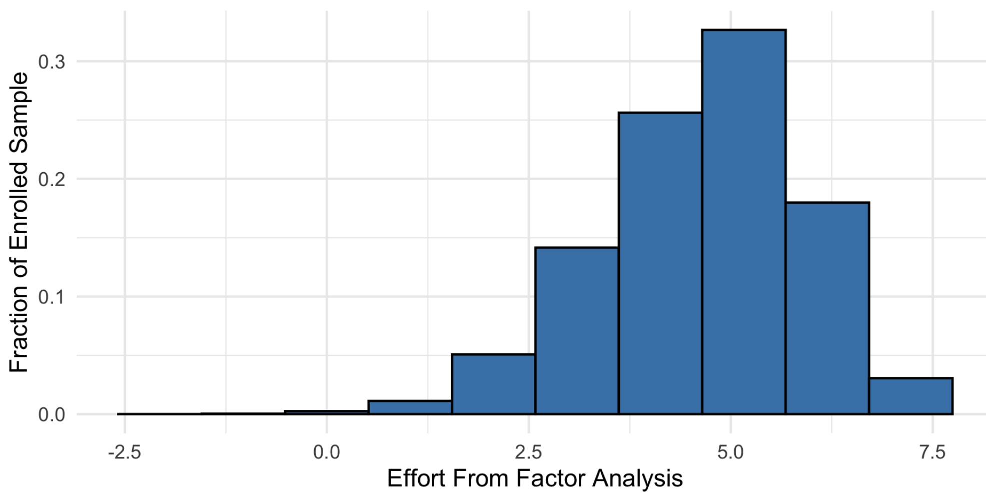


Figure 22: The distribution of latent effort values in the estimation sample.

A.6 Estimation Strategy Details

1. Guess parameters. There are 41 parameters in this version of the model:

    - 13 coefficients for each of the achievement value added equations
    - 3 parameters in the variance-covariance matrix for the value added equations
    - 11 coefficients in the utility equation
    - 1 parameters for the standard deviation of the effort distribution

2. For each student, compute their individual likelihood given the guessed parameters and data:

    - Compute the effort implied by the model (for all options in the student's choice set) using Equation **??**.

- Compute expected math and Spanish scores using effort and Equation **??**.
- For students who enrolled in Grade 7, compute the achievement and effort probabilities. (For students who did not enroll, assign a value of 1 to these probabilities.)
- For all students, compute the multinomial logit probability given in Equation 4.
- Take the product of the three probabilities.

3. Take the log of each individual likelihood, and sum them. Maximize this value with respect to all of these parameters.