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

Gabrielle Vasey†

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 enrolment, study effort, and labor supply, taking into account locally available schooling options and wages. All of these decisions can affect their academic achievement, which is modelled 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 labor laws were enforced, drop out between Grade 6 and Grade 7 would decrease by 6.6%. Expanding the conditional cash transfer, either in terms of the magnitude of the cash benefits or the coverage, while costly, is an operational policy that could achieve similar results. Traveling to school is very costly and policies target accessibility are effective.

JEL Codes: I21, I24, O15.

Keywords: Education Policy, Achievement Gap, Education and Inequality

---

*I would like to thank Jere Behrman, Holger Sieg, Petra Todd, Arthur van Benthem, and Susan Parker for their help and guidance, as well as seminar participants at the University of Pennsylvania, LACEA/LAMES, Canadian Economic Association, Society of Labour Economics, Concordia University Université du Québec à Montréal, Université de Montréal, Dalhousie University, St. Francis Xavier University, and McGill for helpful comments. I am also deeply thankful to Mira Potter-Schwartz, Elizaveta Brover, and Erika Trevino Alvarado for providing excellent research assistance. The Mexican Secretariat of Public Education (SEP) has generously provided data for this analysis. We thank Hector Robles Vaquez for preparing databases and for assistance working with these data. We are grateful for financial support from NSF award #1948943 and from the University of Pennsylvania School of Arts and Sciences internal grants "Making a Difference in Diverse Communities" and the "Dean's Global Inquiries Fund".

†Concordia University; email: gabrielle.vasey@concordia.ca

I. Introduction

When children participate in the labor force, it is often at the expense of their education. Globally, the International Labour Organization estimated that 144 million children under the age of 14 were working in 2012. Unfortunately, much of the progress that was made on this issue in subsequent years was undone by the pandemic. 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. Children who attend school may also work part time and face another choice with respect to the amount of time and effort to dedicate to studying compared to working. Often, laws prohibiting child labor and requiring school enrollment exist, but they are not well enforced. Family socioeconomic status, school availability, school quality, child ability, and earnings opportunities all influence children's time allocation decisions and their resulting academic achievement and attainment.

This paper explores the relationship between child labor, school enrollment and academic achievement in Mexico, and analyzes the impacts of enforcing policies related to labor and education laws. I consider children who 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.

My analysis includes not only the students who go to school full time or work full time, but also the students who combine school and work and the repercussions that working has on their academic achievement. Incorporating this more nuanced choice set is ideal, however this is one of the few empirical studies that includes these choices, which highlights how challenging it is to acquire the required data and setting. The existing studies that do allow for students to combine work and school do not consider the how this impacts academic achievement (Bourguignon, Ferreira, and Leite 2003; Leite, Narayan, and Skoufias 2015). I use individual-level data on school enrollment, test scores, demographics, labor choices, and effort choices to identify the parameters that define the tradeoffs that students are facing.

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. Combined, these data sources create an incredibly rich dataset that has not been used by any studies to date. The administrative data tracks all students in Mexico as they complete national standardized tests and includes information on which students are beneficiaries of the conditional cash transfer Prospera. The data has been used by several recent studies to analyze the impact of Prospera on achievement (Acevedo, Ortega, and Székely 2019; Behrman, Parker, and Todd 2020).

In the model, students decide whether to attend school, and if they attend, they 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, lagged test scores, and working status. I use the model's first-order conditions to solve for an optimal effort level that is specific to each type of school. The data provide five measures of self-reported effort, which I use within a factor model to obtain a single effort index.

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. Achievement is modeled using value-added equations that incorporate student's effort choices. I estimate the model via Maximum Likelihood, where the probability can be decomposed into three conditional probabilities, which each have a closed-form solution.

The identification of the parameters of interest relies mainly on geographic variation of exogenous market conditions and choice sets. To identify the value of school, I use variation in the distances required to travel to school and to identify the value of working, I use variation in local wage offers. Effort is a key mechanism in the model, and the parameters related to the cost and productivity of effort are identified using self-reported effort measures from students.

I find that traveling to a middle school is costly and that students value distance education schools (Telesecundarias) 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 on whether 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 the achievement production function.

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 enforcing the child labor law, which removes the labor option for all children under the age of 14. These estimates provide insight into what fraction of students would choose to enroll if there was no working option and how much achievement would increase if students did not divide their time between work and school. The second counterfactual considers expanding the conditional cash transfer for school attendance, both in terms of benefits amounts and program coverage.

The results of the counterfactual analysis show that enforcing the child labor ban would decrease the drop out rate between Grade 6 and 7 by 6.6%. However, the drop out rate remains high, at just over 7%. Students who are no longer splitting their time between school and work see an increase in their achievement of 4.5% of a standard deviation. When comparing students who were encouraged to enrol versus those who did not, the most striking difference is access to school, with those enrolling having a much closest nearest school. A similar 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.

The paper proceeds as follows. Section 2 lists related literature and the contribution of this paper. Section 3 describes the dataset and setting and provides summary statistics for the variables of interest. Section 4 describes the model of discrete school-work alternatives with endogenous effort choice. 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.

II. Literature

Recently, there have been several papers estimating models of school choice, where schools with differing characteristics are treated as differentiated products (Ferreyra 2007; Epple, Jha, and Sieg 2018; Neilson 2014; Bau 2019; Neilson, Allende, and Gallego 2019). 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 necessary to make the school choice model relevant to developing country contexts where child labor is prevalent.

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.[1] 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 (Todd and Wolpin 2006; Attanasio, Meghir, and Santiago 2011). There also exist some static choice models that include the options of dropping out, enrolling and working part time, or only enrolling (Bourguignon, Ferreira, and Leite 2003; Leite, Narayan, and Skoufias 2015). However, these models also do not examine academic achievement or how working part time affects a child's ability to study. Finally, there are some recent papers that consider the impact of labor on achievement, without incorporating school choice and enrollment decisions. 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 in-

[1] There exists a related literature studying the effects of working in highschool or college, and the effects of this on educational outcomes and human capital accumulation (Stinebrickner and Stinebrickner 2003; Eckstein and Wolpin 1999; Buscha et al. 2012; Le Barbanchon, Ubfal, and Araya 2020).

put in academic achievement. A study using an instrumental variables approach finds 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 (Schultz 2004; Behrman, Sengupta, and Todd 2005; Attanasio, Meghir, and Santiago 2011; Dubois, De Janvry, and Sadoulet 2012). 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 2020).[2] 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 program in Brazil increased academic performance and decreased child labor for beneficiary households.

III. Data and Setting

The data requirements for answering research questions related to child labor, school enrollment, and academic achievement are high. Among other variables, the labor choice, school choice, and achievement realization for each student must be observed. The data set that I use provides the above mentioned variables and more. In addition, there is quasi-random variation, since each primary school has a different choice set of middle schools, as well as different labor market conditions. Unfortunately, there are some variables that are not available in this setting, most notably information on parental wages. The absence of these data will inform the modeling choices that I make in the next section.

Not only does the data set provide the majority of the required variables, but

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

Mexico as a country is an ideal setting to study this question. In 2010, the year analyzed in this paper, Mexico had Education Regulation that defined Grades 1 through 9 as compulsory, and Labor Regulation that prohibited labor of minors under the age of 14. However, 7.9% of children age 12 and 13 reported not being enrolled in school in the 2010 Census. Further, over a quarter of students in Grade 7 reported working at least one day a week in a national survey. In addition, increasing student enrollment has been a target for the Mexican government for decades, with programs such as the conditional cash transfer and the distance education schools. These ensure that the majority of students, even in rural areas, have access to a local school.

A. Data Sources

To carry out this research, I use a newly available merged dataset. This dataset is comprised of several components which come from two main 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, Ortega, and Székely 2019; Behrman, Parker, and Todd 2020). The third 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 the previously mentioned source. The 2010 Census is used to access information on children between the ages of 12 and 18, 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, I have their national standardized test scores, their school IDs (with associated school information), individual demographics, household demographics (including conditional cash transfer status), and the average municipal wage conditional on age, gender, family background and school attendance.

B. Estimation Sample

The analysis in this paper focuses on students across Mexico in Grade 6 in 2008 who progress to Grade 7 in 2009.[3] 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.[4] 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. I assume that these students have dropped out of school.

In Grade 7 in 2009, there are 107,898 students for whom we have survey answers from themselves and their parents.[5] The mean age of the students in Grade 7 is 13, with a minimum age of 12 and a maximum age of 17. 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.

[3] There are three states (out of 32) 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.

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

[5] 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.

C. Key Variables

C..1 Test Scores

Standardized test scores in math and Spanish are used as a measure of student achievement. The test administrators (SEP) standardized the tests in the base year, 2008, to have a mean of 500 and a standard deviation of 100. The same transformation was used in subsequent years.[6] 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.

Table 1
Summary Statistics for Test Scores

	Mean	Standard Deviation
Grade 6 Math	525.7	120.8
Grade 6 Spanish	518.0	106.3
Grade 7 Math	504.3	98.7
Grade 7 Spanish	502.1	98.1

Scores are for the estimation cohort used throughout the paper (64,215 students have both Grade 6 and 7 test scores and 5167 students have only Grade 6 scores). The distributions of the test scores are approximately normal, as shown in Appendix A..

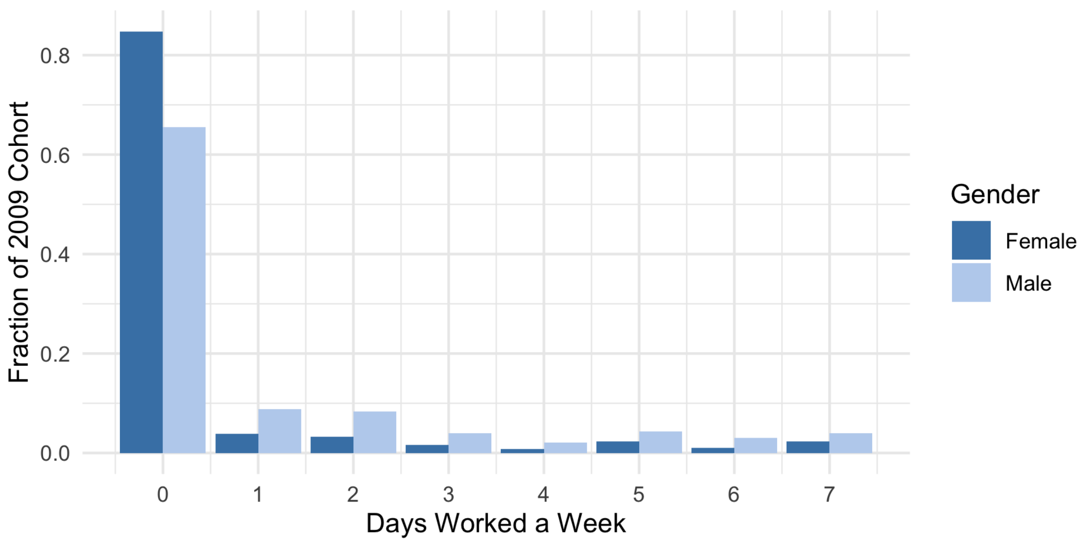
C..2 Labor Decision

To observe the labor decision of the students, I use a question from the student survey which asks: “On average, how many days a week do you work?". Figure 1 shows the responses, divided by gender. Boys work more than girls, and the majority of students are not working. 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

[6] This is not equivalent to standardizing the scores each year, as is apparent from the means presented.

14 and older the mean is 1.68, so older children are working substantially more than the younger children.

Figure 1
Days Worked per Week



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

Although I will not be considering different occupation types in this project, it is of interest to know what kinds of labor children were engaging in during this time period in Mexico. From the Census, the most common occupation type for 12 year olds was agriculture (maize, beans, livestock, flowers, vegetables, fruits) with the next most common being a sales worker or working in a store. Other occupations reported included street vendors, food preparation and a support worker for construction. In the student survey, there is a question inquiring about the reasons for working, and 59% of students reported working for their family.

C..3 Student Effort

Achieving a high test score and earning a wage at a job both take time and energy. To capture this, and to understand how combining school and work may impact achievement, I incorporate effort into my analysis. The rich data set provides five self-reported measures related to effort. The questions 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.

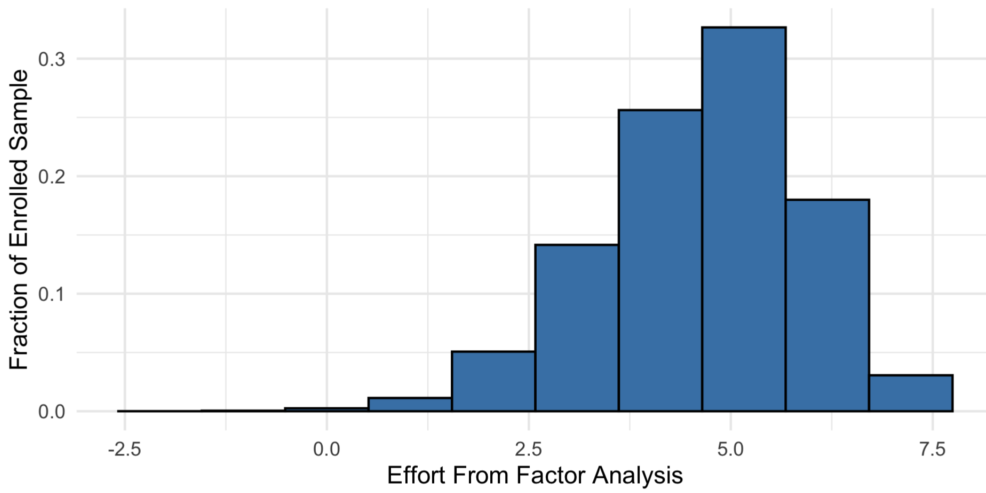
The first measure, the number of hours studied per day, is cardinal. The other four measures are ordinal variables, as they are answered on a Likert scale. To combine them into one value, I use factor analysis. This analysis is done outside of the model estimation, and uses polychoric correlations to 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, $\tilde{e}_{ijL}^{M}$, which combines the information from the student's responses to the five effort questions. Figure 2 presents a histogram of the new continuous effort variable. The effort values are almost all positive and the distribution appears to be approximately normal.[7] Estimation details and results are in Appendix B.

C..4 Wages

It is necessary to know what wage each of the children could be earning if they decided to work. Unfortunately, wages are not included in the survey data, so I impute wage offers for all students using census data. The census contains the working status, enrollment status, and the monthly wages earned for children across Mexico. Other important information such as the age and gender of the child, the education level of their parents, and the municipality in which they live is also recorded.

[7]Negative values of effort are possible, though rare, since the last two effort questions have negative loading factors.

Figure 2
Latent Effort Variable Distribution



The distribution of latent effort values in the estimation sample. This variable is combines information from five questions related to study effort using factor analysis.

To account for non-random selection into working, a two-step Heckman selection model is estimated. The first step involves estimating a probit model where the outcome is the probability of working. 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. The second step is a linear regression, which incorporates a control term created using estimates from the first step, and interaction terms between all of the other covariates. Regressions are estimated separately for girls and boys. For details on the wage estimation and parameter estimates, see Appendix F..

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

Table 2 contains results from the imputations. The results are divided by gender of the child, and by the school enrollment status. The mean and standard deviation are shown for the monthly wage. The monetary values are in 2010 pesos. For children working full time, the imputed wages for females is 1654 pesos per month and for males it is 867 pesos per month. The part time wages for students who are enrolled in school are significantly lower, at 867 pesos per month for girls and 887 pesos per month for boys.

Table 2
Summary Statistics for the Monthly Wage Imputations

		Female	Male
Work and School	Mean	867	887
	Standard Deviation	(393)	(353)
Only Work	Monthly Wage	1654	1717
	Standard Deviation	(411)	(339)

Summary statistics from the wage regressions, which were estimated separately by gender. Monetary values are in 2010 pesos. The sample of children used for the imputation is the estimation sample used throughout the paper.

C..5 School Types

There are four different types of middle schools in Mexico: General, Technical, Telesecundarias, and Private. Technical middle schools have a focus on vocational studies. Telesecundarias, 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 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 3 contains summary statistics for the four different types of schools in Mexico. From the table it is apparent that there are many small telesecundarias, predominately in rural areas. 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 telesecundaria 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 3
Summary Statistics for School Types

	General	Technical	Telesecundaria
Number of Schools	537	357	856
Number of Students	30,984	20,168	13,063
Proportion of Cohort	0.483	0.314	0.203
Proportion Female	0.519	0.514	0.496
Proportion CCT	0.141	0.182	0.696
Proportion Rural	0.367	0.406	0.748
Mean School Cohort Size	86.3	85.5	18.7

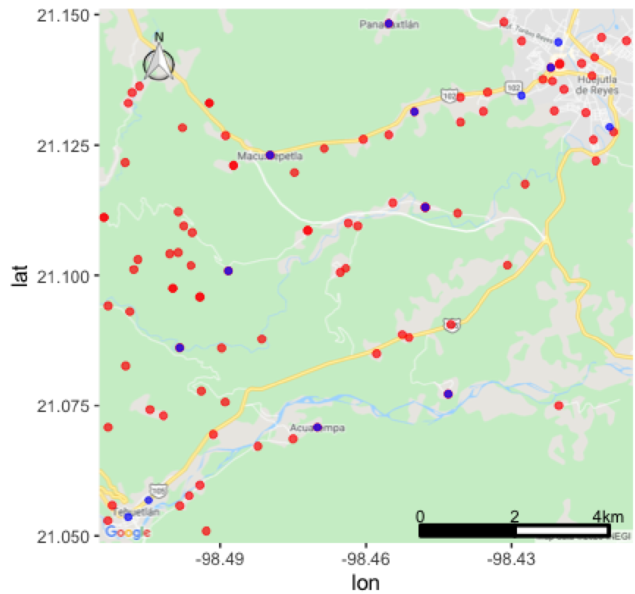
Summary statistics for the three types of middle schools (private schools are not included in the analysis, but less than 8% Grade 7 students attended them in 2009). This table is created using the estimation sample (the 64,215 students who enrolled in Grade 7).

C..6 Distances and 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 3 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 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 4. It is also possible to calculate distance using roads and paths on Google Maps, but this does

Figure 3
Map with Schools in Example Neighborhood



Map of all primary schools (red) and middle schools (blue) in a rural region of Mexico. The upper right corner contains a city while the remaining region is considered to be rural. There are many more primary schools than there are middle schools available.

not capture many of the rural pathways.

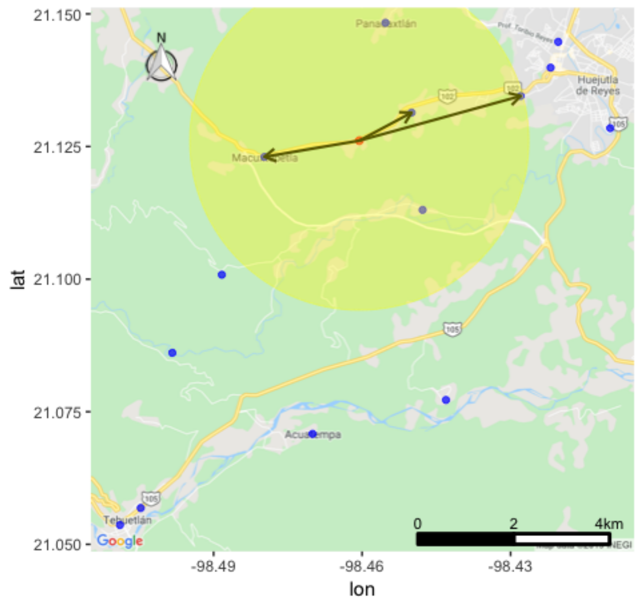
For the estimation, I have to define which middle schools each student considers when making their school choice. To do this, I create a circle around the primary school and consider all middle schools within the circle to be in the choice set, as shown in Figure 4. However, choosing the same radius 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.[8]

D. *Data Patterns*

To quantify the impact of enforcing child labor laws on educational attainment and achievement, it is crucial to understand the relationships between working,

[8] 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 4
Map Showing Choice Set in Example Neighborhood



Map of a primary school (red) with the middle schools (blue) included in its choice set. The choice set is comprised of all schools included in the yellow circle. The arrows represent the actual choices of students from the primary school. There are two schools that were not chosen by students in the primary school, but are included in the choice set given their geographic proximity.

study effort, achievement, and the many other inputs from the setting. The following subsections describe these patterns and correlations in the data. This is helpful in understanding the research question, and also informative for modelling. Finally, the estimated model should be able to reproduce these patterns and correlations, which is confirmed in the Results section.

D..1 Working and Achievement

When discussing working while in school, a main concern is that the test scores of students who are working could suffer. The simple regression shown in Table 4 does show a significant negative relationship between working and Grade 7 test scores. The covariate "Working" is a dummy variable, and in column (1) it is equal to 1 for all students who report working at least 1 day a week, in column (2) it is 1 for all students who report working at least 2 days a week, and so on. This is a descriptive regression, so the results should not be interpreted as causal, however it does control for the student's lagged test scores, their gender and age, their parental education levels, and whether they are a beneficiary of the conditional cash transfer. The magnitude of the relationship increases as the number of days a week working increases. Students who work at least 3 days a week are found to have 5% of a standard deviation lower test scores than their peers. This magnitude is equivalent to the increase in test scores that students have if their mother has at least a middle school education.

D..2 Working and Study Effort

Although the relationship between working and achievement can be studied directly, it is more informative to investigate the underlying mechanisms. The main mechanism that comes to mind that connects both work and academic achievement is time and energy. For this project, I will bundle them together, and call the overall measure study effort (the variable is described in detail in Section C..3). If students are working, they have another use of their time. Table 5 shows that there is a negative correlation between working at least one day a week and the effort variable. The correlation decreases in magnitude as more controls are added, however the significance remains. In the final column, students who work at least one day a week have 5% of a standard deviation lower effort values than students who are not working.

D..3 Study Effort and Achievement

For study effort to be a valid mechanism between working and achievement, there must also be a correlation between effort and achievement in the data. Figure 5 shows that higher test scores in both math and Spanish are correlated with higher values of effort (without controlling for any covariates). While controlling for variables, such as lagged test scores and parental education, does decrease the magnitude of the relationship, Table 6 shows that the significant positive relationship still exists. These results also provide evidence that the effort measure created in this paper is picking up an important input into test scores, and that this input is not captured by lagged test scores and other demographic variables.

D..4 School Accessibility and Enrollment

When deciding whether to enroll in school or not, students take into consideration both the availability of schools and their outside option of working. The farther

Table 4
Days Worked per Week and Test Scores

	Dependent variable:			
	Grade 7 Test Score			
	(> 0 Day)	(> 1 Days)	(> 2 Days)	(> 3 Days)
	(1)	(2)	(3)	(4)
Working	-0.047***	-0.070***	-0.106***	-0.150***
	(0.012)	(0.013)	(0.015)	(0.017)
Lagged Tests Scores	0.675***	0.675***	0.674***	0.674***
	(0.003)	(0.003)	(0.003)	(0.003)
Female	0.183***	0.182***	0.183***	0.183***
	(0.010)	(0.010)	(0.010)	(0.010)
Age	-0.078***	-0.078***	-0.077***	-0.077***
	(0.008)	(0.008)	(0.008)	(0.008)
Mom Middle School	0.041***	0.041***	0.041***	0.041***
	(0.011)	(0.011)	(0.011)	(0.011)
Dad Middle School	0.072***	0.072***	0.072***	0.072***
	(0.012)	(0.012)	(0.012)	(0.012)
Prospera	0.129***	0.130***	0.130***	0.131***
	(0.012)	(0.012)	(0.012)	(0.012)
Constant	3.890***	3.890***	3.889***	3.886***
	(0.103)	(0.103)	(0.103)	(0.103)
Observations	64,356	64,356	64,356	64,356
R^2	0.505	0.505	0.505	0.506

Note: *p<0.1; **p<0.05; ***p<0.01

Correlation between days worked per week and test scores in the data. The covariate “Working" is a dummy variable, and its definition changes depending on the column. In column (1), “Working" is equal to 1 if students work at least 1 day a week. In column (2) “Working" is equal to 1 if students work at least 2 days a week, and so on. The dependent variable is the sum of each student’s Grade 7 math and Spanish test scores.

Table 5
Working and Study Effort

	Dependent variable:		
	Effort		
	(1)	(2)	(3)
Working	-0.204***	-0.086***	-0.070***
	(0.012)	(0.011)	(0.012)
Lagged Tests Scores		0.172***	0.171***
		(0.003)	(0.003)
Female			0.128***
			(0.010)
Age			-0.121***
			(0.008)
Mom Middle School			-0.015
			(0.011)
Dad Middle School			0.062***
			(0.011)
Prospera			0.234***
			(0.012)
Constant	4.698***	2.911***	4.215***
	(0.006)	(0.028)	(0.100)
Observations	64,356	64,356	64,356
R^2	0.005	0.065	0.077

Note: *p<0.1; **p<0.05; ***p<0.01

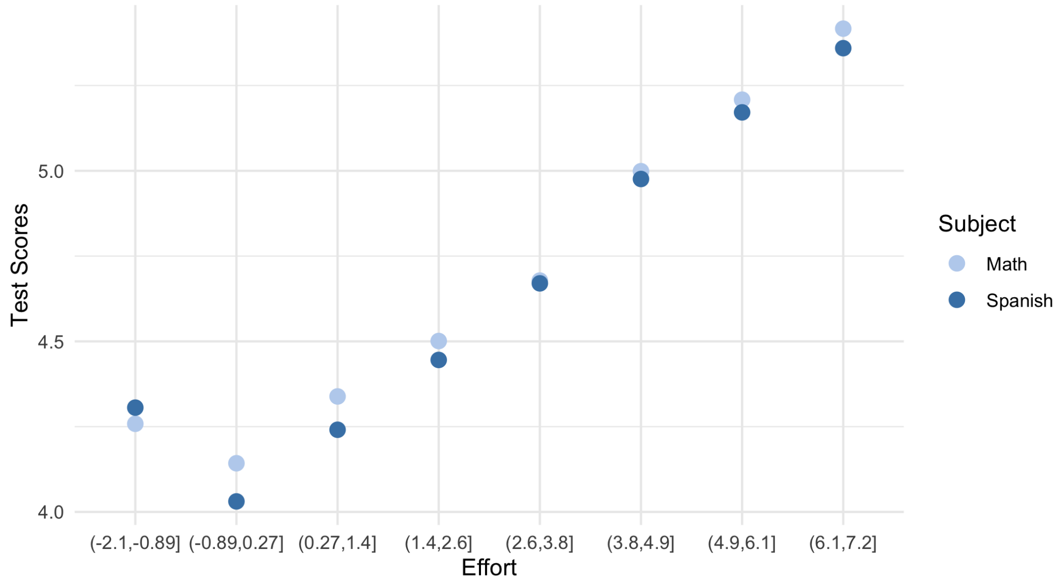
Correlation between days worked per week and study effort in the data. The variable "Working" is a dummy variable equal to 1 if a child reports working at least 1 day a week. The effort variable is the continuous variable created with factor analysis.

Table 6
Study Effort and Achievement

	Dependent variable:		
	Grade 7 Test Score		
	(1)	(2)	(3)
Study Effort	0.430***	0.190***	0.182***
	(0.005)	(0.004)	(0.004)
Lagged Tests Scores		0.655***	0.645***
		(0.003)	(0.003)
Female			0.166***
			(0.010)
Age			-0.057***
			(0.008)
Mom Middle School			0.045***
			(0.011)
Dad Middle School			0.062***
			(0.011)
Prospera			0.082***
			(0.012)
Constant	8.026***	2.460***	3.116***
	(0.025)	(0.030)	(0.102)
Observations	64,356	64,356	64,356
R^2	0.094	0.517	0.521
Note:	*p<0.1; **p<0.05; ***p<0.01		

Correlation between study effort and test scores in the data. Test scores are the sum of Grade 7 math and Spanish scores. The effort variable is the continuous variable created by factor analysis.

Figure 5
Study Effort and Test Scores



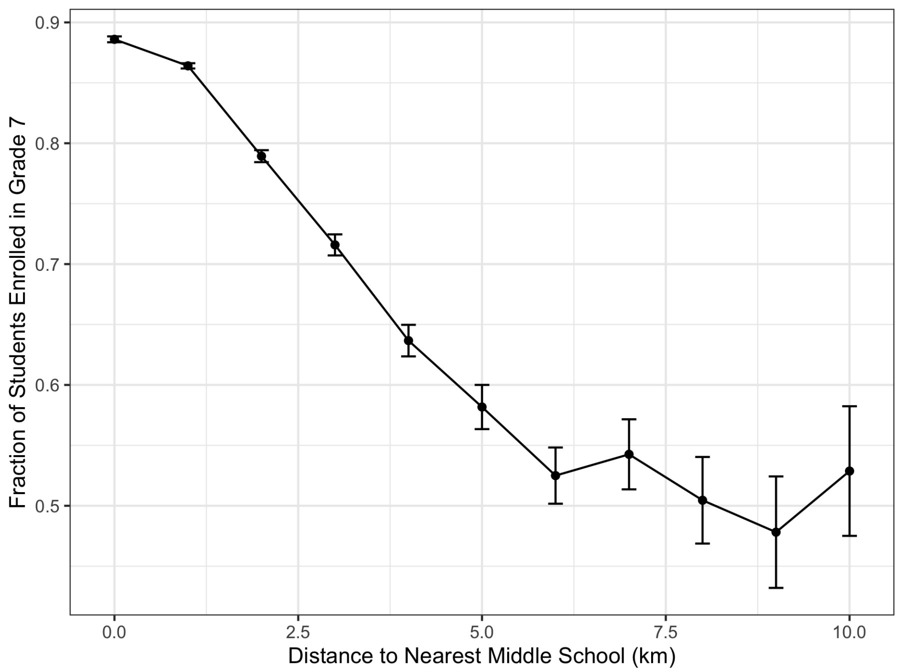
The correlation between study effort and Grade 7 test scores. The effort variable on the x-axis is binned into eight categories, and the mean test scores for students with effort values in the corresponding bin are calculated. There is a positive relationship between effort and test scores.

away a middle school is from their primary school, the higher the cost of traveling there. Figure 6 shows that students who have no schools in their area are more likely to drop out than the students who have middle schools nearby.

D..5 Wages and Working

The imputed wages depend not only on student characteristics, but also based on the municipality in which the student resides. This provides geographic heterogeneity in the wages. However, this geographic variation is not useful if it is not correlated with the choices that students are making. Table 7 contains estimates from a linear probability regression, where the observations are at the municipality level. For each municipality, the average monthly wage is calculated, as well as the average lagged test scores, the average parental education level, and the mean proportion of students receiving the conditional cash transfer. Wages are normalized in this regression so that a standard deviation is equal to 1. The results show that increasing the mean wage by one standard deviation is correlated with an increase of almost 1 percentage point in

Figure 6
Enrollment Rates and Access to Middle Schools



The relationship between enrollment rates and middle school accessibility. Each primary school is categorized by the distance to the nearest middle school, which is the x-axis. For each primary school, the fraction of its students who continue to Grade 7 is also calculated. The graph shows that primary schools that have middle schools near by have a higher fraction of their students enroll in middle school.

the dropout rate of the municipality. This is very significant given that the national average dropout rate is less than 8%.

IV. Model

The model captures the different choices that students make as they progress from primary school (Grade 6) to middle school (Grade 7). They choose 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 telesecundarias (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

Table 7
Dropout Rates and Local Wages

	Dependent variable:
	DropRate
Full Time Wage	0.008***
	(0.002)
Lagged Tests Scores	-0.023***
	(0.004)
Mom Middle School	0.051***
	(0.011)
Dad Middle School	-0.024***
	(0.008)
Prospera	-0.030***
	(0.008)
Constant	0.306***
	(0.045)
Observations	936
R^2	0.222
Note:	*p<0.1; **p<0.05; ***p<0.01

Correlation between the dropout rate and the mean wage in a municipality. The observations are at the municipality level. For each municipality, the average wage is calculated, as well as the average lagged test scores, the average parental education level, and the mean proportion of students receiving the conditional cash transfer. Wages are normalized in this regression so that a standard deviation is equal to 1.

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.

A. 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.[9] 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 with labor choice L is given by

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

[9] In the ideal scenario a family budget constraint would be included. Unfortunately, although the survey does contain a question on family income, the responses are in very coarse bins, and are not fine enough to include in a budget constraint. In the results section, I discuss how not including a budget constraint could bias the results.

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, $\hat{A}^7_{ij}(e_{ijL})$ depends on student characteristics, middle-school characteristics, and students' effort choices e_{ijL}. Students may care differently about their score 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 telesecundaria, Technical or General. Students potential distaste for working while in school is captured by the coefficient α_8.

A linear and quadratic term for effort are included in the student utility. This allows for flexibility and also ensures a solution for the optimal effort for each student. 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.

B. *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 (details are in Section C..4). Wages are allowed to depend on age, gender, school attendance, parental education, and geographic location (either urban/rural and municipality).

C. *Expected Test Scores*

For students who choose to enroll in school, their test score is 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.

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

The last term, $e_{ijL}Z_j$, is an interaction between the student's effort level and the school type, allowing for effort to be more or less productive depending on the type of school attended. The test score is the sum of the Grade 7 math and Spanish test scores. Students are assumed to not know the error term 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.

D. 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:

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

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

$$e^*_{ijL} = \frac{-\left((\alpha_3 + \alpha_4 PEduc_i + \alpha_5 \mathbb{1}\{CCT > 0\})(\delta_3 + \delta_6 Z_j) + \alpha_{i,9} + \mathbb{1}\{L = PT\}\alpha_{10}\right)}{2\alpha_{11}} \tag{2}$$

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 ($\hat{A}^7_{ije^*}$).

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

The final result is that each student has an optimal school j and labor option L, and an optimal effort given these choices, e^*.

V. Estimation

Model parameters are estimated using Maximum Likelihood. Define

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

as the joint probability of choosing school j, labor option L, having Grade 7 test score A_{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_{ij} is normal.

Define $D_{ijL} = 1$ if student i chose school j and labor option L. The likelihood is then,

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

The joint probability can be decomposed into the product of conditional probabilities. The variable $\tilde{e}^M_{ijL}$ is the effort variable in the data. Two of the conditional probabilities depend on e^*_{ijL} and 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. Conditioning variables in probabilities are dropped in the probability expressions if the probability does not depend on them.

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

Consider each of the three probabilities in the likelihood. The first is the probability of observing the Grade 7 test score:

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

The errors for the achievement production function are distributed iid 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.

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

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.

To calculate the standard errors, I estimate a sandwich-type covariance matrix. Details are in Appendix E..

A. Identification

There are 30 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: $\{\delta_k\}_{k=1}^{11}$, σ_T
- Effort: σ_E

There are 12 parameters associated with achievement. They are estimated with a value added equation. Each student who attended Grade 7 has a test score in both Math and Spanish, and the sum of these scores is used as the outcome variable. Each student also has lagged test scores in both subjects, as well as data on the 10 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). School types 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, telesecundaria). 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.

VI. Results

The utility function parameter estimates are shown in Table 8 and the test score production function parameters are shown in Table 9. 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.

The estimate for the cost of dropping out of school in the utility function is large, and increases further if parents have middle-school education. Technical schools are estimated to be valued higher than general schools and telesecundarias are estimated to be perceived significantly worse than the other two school types.

The average expected test score has a large, 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 test scores by just over 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 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.

The coefficient estimates in the achievement production function, shown in Table 9, are fairly intuitive. Lagged test scores have a significant impact, with the coefficient on lagged Spanish being larger than that on lagged math. Female students and those with higher ages are estimated to do worse. The value added of a Technical schools is essentially the same as a General school whereas telesecundarias are estimated

Table 8
Coefficient estimates for parameters in the utility function.

Coefficients	Estimates	Std.Error
Distance	-45.249	4.006
Distance Squared	1.842	0.163
Dropping Out	-31.997	13.87
Dropping X Parent Educ	-9.819	5.156
Technical School	15.979	1.462
Telesecondary School	-45.242	3.261
Expected Score	26.829	1.502
Expected X Parent Educ	-0.029	2.436
Expected X CCT	0.065	0.023
Working Part Time	-54.89	4.053
Linear Effort	-35.934	3.259
Effort X Lagged Score	0.342	0.092
Effort X Female	0.317	0.09
Effort X Parent Educ	0.162	3.833
Effort X Work	-0.166	0.047
Quadratic Effort	-1.101	0.294
Random Coef St Err	1.228	0.004
Scale Parameter	41.681	3.685

to be worse, although the coefficient estimate is not significant. The school quality coefficient is negative. Finally, effort has a large positive coefficient and it is precisely estimated. Effort is estimated to be slightly more productive in telesecundaria schools compared to General and Technical schools, however the change in productivity is small in magnitude.

A. Model Fit

The following tables and figures show the fit of simulated data from the model compared to the true data. The estimation targets three main outcomes: test score,

Table 9

Coefficient estimates for parameters in the achievement functions.

Coefficients	Achievement_Estimates	Std.Error
Intercept	1.684	0.378
Lagged Math	0.266	0.018
Lagged Spanish	0.433	0.02
Female	-0.08	0.023
Age	-0.191	0.01
Technical School	0.006	0.026
Telesecondary School	-0.068	0.071
School Quality	-0.115	0.011
Effort	1.574	0.109
Effort X Technical	-0.004	0.001
Effort X Tele	0.019	0.005
Residual Standard Error	1.25	0.004

effort, and school choice. Table 10 reports the means (and standard deviations when appropriate) for these main outcomes. The overall fit is very good. The exception is that the test score distribution is slightly off, as shown in Figure 7. The simulated effort distribution, shown in Figure 8, has a much closer fit to the data.

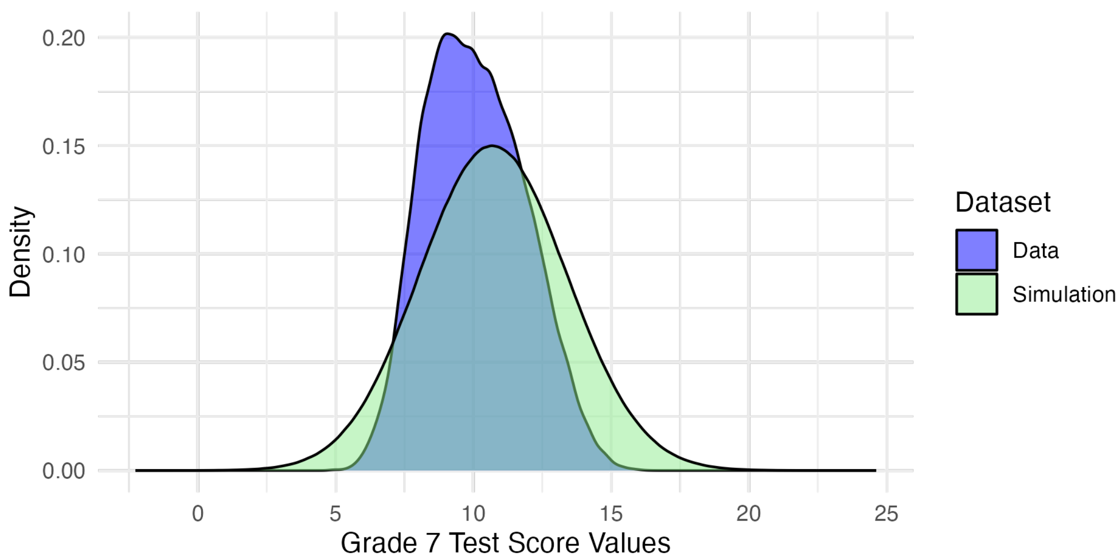
Lagged test scores are assumed to be exogenous in the model, and are a proxy for ability. In the data, this variable is highly correlated with both effort and drop out, two endogenous choices in the model. Figures 9 and 10 show that these relationships are almost perfectly captured by the model. Specifically, the model is able to recreate the non-linear relationshihp between lagged test scores and effort.

Focusing on the relationship between school accessibility and dropout, Figure 11 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 U-shaped pattern matches, but it is apparent that the model is overestimating dropout rates for students who have a middle school in the same location as their primary school, or who have a middle school very far away.

Table 10
Model fit for relevant means and standard deviations.

Outcome Variable	True Mean	Simulated Mean	True St.Dev.	Simulated St.Dev.
Test 7	10.07	10.73	1.79	2.65
Effort	4.65	4.65	1.28	1.28
Fraction Drop	0.07	0.07		
Fraction General	0.45	0.45		
Fraction Technical	0.29	0.29		
Fraction Telesecondary	0.19	0.19		
Fraction Work PT	0.24	0.22		

Figure 7
Goodness of Fit: Test Score Distribution

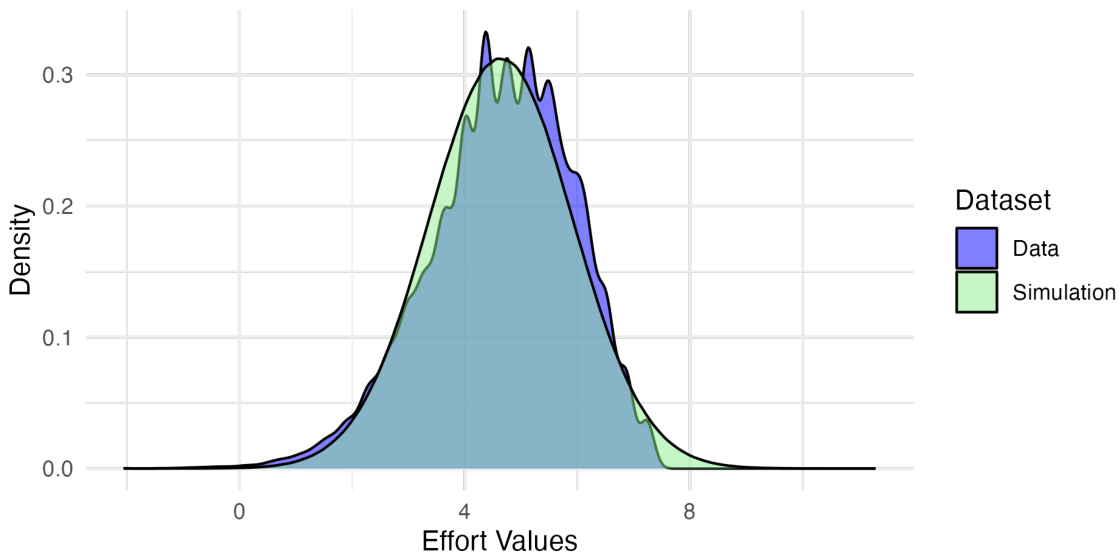


This figure shows the goodness of fit with respect to the Grade 7 test score distribution.

VII. 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 first policy I consider is an enforcement of the child labor laws, which would prohibit all chil-

Figure 8
Goodness of Fit: Effort Distribution



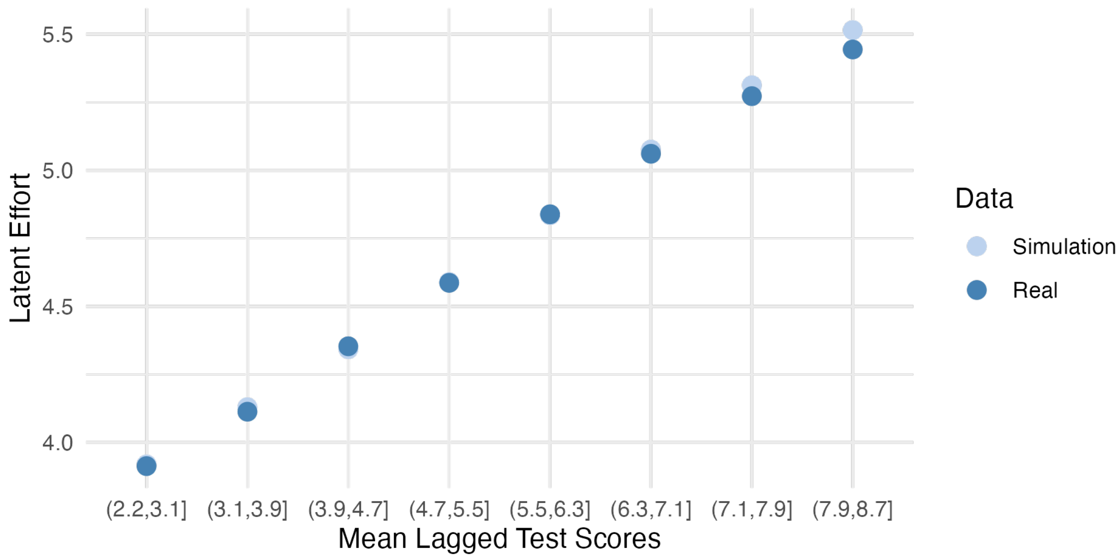
This figure shows the goodness of fit with respect to the effort distribution.

dren younger than 14 from working. Given that this policy would be very costly to implement, and possibly even impossible, I then consider a set of alternative policies that would lower the dropout rate by a similar amount. I find that conditional cash transfers, which are already used widely in Mexico, can be effective, but that they must increase in magnitude or better target at risk students. Access to school is a primary issue, and another set of successful counterfactual policies consider the impact of minimizing the cost of traveling.

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

The first counterfactual involves removing all of the labor options. Anyone who chose to work originally, either full-time or part-time, is affected by this policy. Stu-

Figure 9
Goodness of Fit: Lagged Achievement and Effort



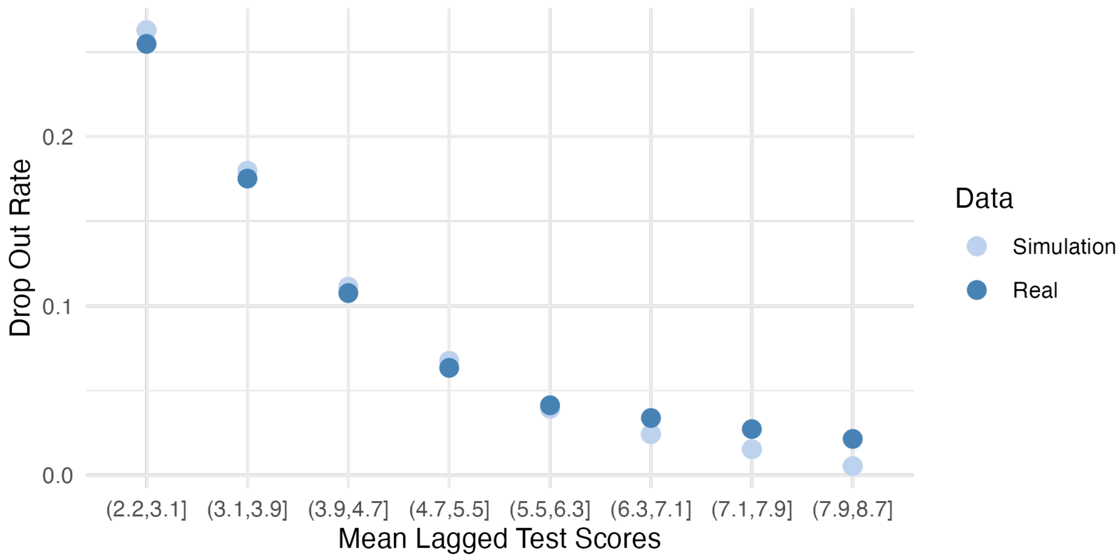
This figure shows the goodness of fit of the model with respect to the relationship between lagged achievement and effort. The effort variable from the data and the effort variable generated by the model are plotted as a function of average lagged test scores.

dents who had chosen to drop out must decide if they will enroll in school given the absence of a full-time wage, or if the cost of school is still too high. Students who were working part-time will stay enrolled in school, as the value of their outside option has decreased more than the value of enrolling in school without work. The results of the policy are shown in Table 11. The drop out rate decreases by 6.7%, but still remains above 7%. Mean effort and test scores increase slightly.

In terms of effort and achievement, banning labor has a large effect on the 22% of students who were working part-time. Table 12 shows that for this group, effort increases by a more meaningful 5.9% of a standard deviation, and test scores increase by 4.4% of a standard deviation.

The change in drop out rate from this policy is not substantial. To better understand why, it is important to see how the students who chose to enroll differ from those who stayed out of school. Table 13 shows how the students differ by background characteristics. Overall, students who enrolled were slightly less disadvantaged than those

Figure 10
Goodness of Fit: Lagged Achievement and Drop Out



This figure shows the goodness of fit of the model with respect to the relationship between lagged achievement and dropping out.

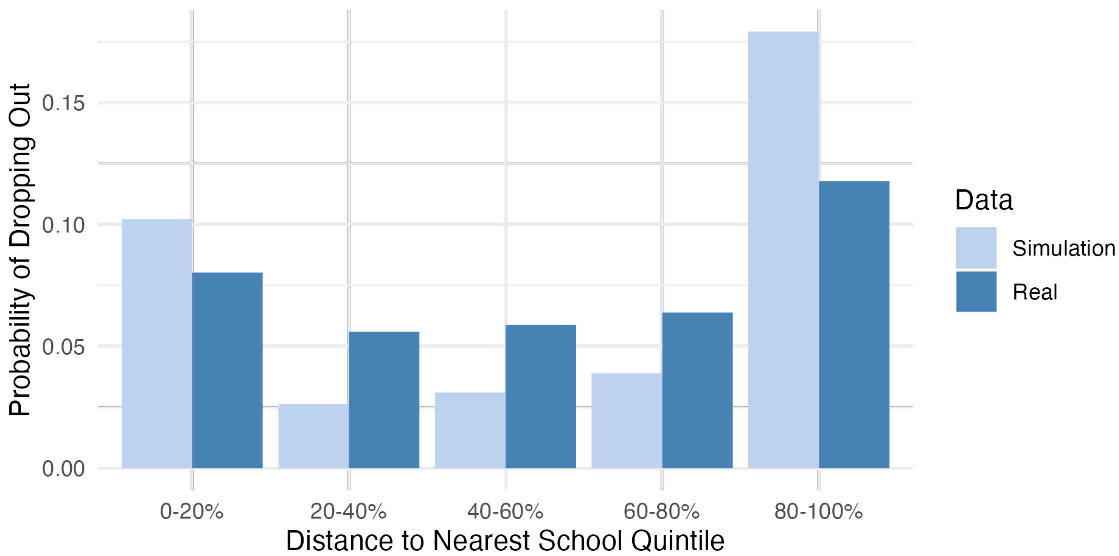
Table 11
Changes in outcomes from Counterfactual 1.

Outcome Variable	Estimated Model	Counterfactual	Percent Change
Fraction Work PT	0.22	0.00	-100.00
Fraction Drop	0.07	0.07	-6.67
Mean Effort	4.65	4.67	0.37
Mean Test	10.73	10.75	0.20

who did not enroll. However, the most striking difference between the two groups is in the access to school. Those who chose to enroll had a nearest school that was almost 1 kilometer closer than those who did not enroll. This underlines the importance of school access to this problem.

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

Figure 11
Goodness of Fit: Enrollment and Accessibility



This figure shows the goodness of fit with respect to the enrollment rates and the school accessibility. The fraction of students who dropout is broken down by how far away the nearest school is from their primary school.

Table 12
Changes in outcomes from Counterfactual 1 for students who would like to work part time, but stayed enrolled when they could not.

Outcome Variable	Estimated Model	Counterfactual	Change in SD (%)
Effort	4.59	4.67	5.89
Test	10.69	10.80	4.44

well established policy, the conditional cash transfer, that could be modified. In my next counterfactual, I consider changing the values of the conditional cash transfer and expanding eligibility for the program.

Figure 12 shows the reduction in dropout rates for four different conditional cash transfer policies. The x-axis shows an increase in the amount of the transfer, ranging from the 2010 Prospera transfer amount, to 9 times the transfer amount. The y-axis shows the dropout rate, from 5% to 8%. Three horizontal lines are marked on the graph. The first, labeled "Simulation" is the dropout rate in the baseline simulation.

Table 13

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

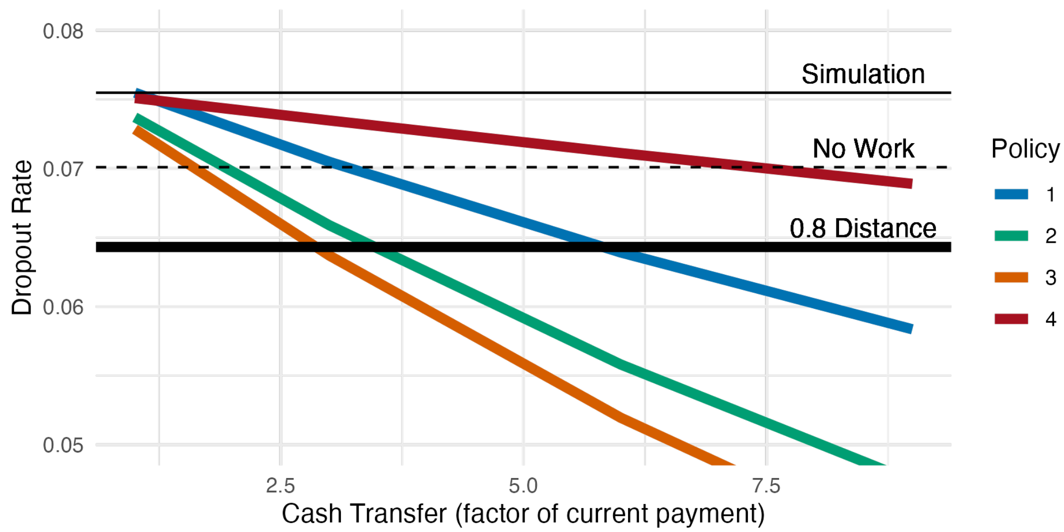
Background Variables	Stayed In School	Dropped Out
Lagged Math	4.66	4.57
Lagged Spanish	4.60	4.50
Female	0.48	0.49
CCT	0.50	0.59
Age	12.35	12.35
Mom has Middle School	0.28	0.24
Dad has Middle School	0.29	0.25
Family Income above Mean	0.23	0.20
Nearest School	1.06	1.90

The dashed-line below it, labeled "No Work" is the dropout rate when the labor laws are enforced and no work options are available. Finally, the bold line labeled "0.8 Distance" is a counterfactual simulation that reduces the distance for all students by 20%. This line illustrates the cost of travel time for students, and how it is a main factor driving the high dropout rate.

The four policies in Figure 12 change who is offered the conditional cash transfer. Policy 1 considers increasing the transfers to the current beneficiaries, which would be very simple to implement. Policy 2 extends the transfer beneficiaries to those who currently received Prospera, and those who have an income below the median. Policy 3 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. Finally, Policy 4 offers the transfer to all students whose nearest middle school is more than 2km from their primary school.

The results show that increasing the conditional cash transfer payment is a very effective way of decreasing the dropout rate, but that it would come at a very high cost. 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. When considering the costs of these policies,

Figure 12
Counterfactual Policy: Dropout Rate and Conditional Cash Transfers



The fraction of students who dropout when considering four different conditional cash transfer policies. Policy 1 offers the transfer to the actual beneficiaries in the data. Policy 2 offers the transfer to actual beneficiaries and students whose family earns below the median income. Policy 3 offers the transfer to all students who would choose to drop out without a transfer. Policy 4 offers the transfer to all students whose nearest middle school is more than 2km from their primary school.

the figure does not tell the full story, because the number of beneficiaries is just as important as the transfer amount. Table 14 contains information on the number of beneficiaries for each policy, and the transfer amount and total cost to reduce dropout to the "No Work" level. Policy 3, which is not operational, is the least cost option. After this, it is actually Policy 4, which is based on school access, that reduces dropout for the least cost. This highlights the importance of targeting, and suggests that programs such as school buses could be highly effective if offered in remote areas.

Students who are most at risk of dropping out are predominantly in rural areas, and their nearest school option is often a telesecundaria. The parameter estimates reveal that there is a negative perception of these schools. The final counterfactual policy considers the impact of a change in perceptions, in which telesecundarias were perceived as having the same value as general schools. Table 15 shows that there are

Table 14
Comparing costs for different CCT policies. The transfer amounts are in factors of the 2010 transfers. The total costs are also in factors of the actual 2010 total cost.

Policy	Beneficiaries	Students Impacted	Transfer Amount	Total Cost
1	Beneficiaries in data	18,576	3.18	3.18
2	Beneficiaries + income below median	43,994	1.92	4.56
3	Any student who drops	5,178	1.60	1.45
4	Nearest school more than 2km	3,667	7.38	2.46

substantial gains to be realized from this simple change in perceptions. The dropout rate would decrease to 5% and the fraction of students who are enrolled in telesecundarias would increase from 19% to 26%. Test scores, effort, and the fraction working in school remain essentially unchanged.

Table 15
Changes in outcomes if Telesecundarias were valued the same as General schools.

Outcome Variable	True Mean	Simulated Mean	Counterfactual Mean
Test 7	10.07	10.73	10.73
Effort	4.65	4.65	4.66
Fraction Drop	0.07	0.07	0.05
Fraction General	0.45	0.45	0.41
Fraction Technical	0.29	0.29	0.27
Fraction Telesecondary	0.19	0.19	0.26
Fraction Work PT	0.24	0.22	0.23

VIII. 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 labor option from student's choice sets, I evaluate the impact of that working has on both enrollment and achievement. Achievement and effort increase when students are not splitting their time. The dropout rate decreases by 6.6%, but still remains high. This policy would also be challenging to implement. I evaluate a second policy, which is to increase the magnitude or the number of beneficiaries of the conditional cash transfer. I find that with more money and better targetting, a similar decrease in drop out can be achieved by incentivizing students to enroll, instead of banning them from working.

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 (telesecundarias). In ongoing work, I am considering a range of such policies.

References

Acevedo, Ivonne, Araceli Ortega, and Miguel Székely. 2019. "Rendimiento Escolar y Transiciones Laborales con Transferencias Condicionadas en México." *Working Paper*.

Attanasio, Orazio P, Costas Meghir, and Ana Santiago. 2011. "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progresa." *The Review of Economic Studies* 79 (1): 37–66.

Bau, Natalie. 2019. "Estimating an Equilibrium Model of Horizontal Competition in Education." *Working Paper*.

Behrman, Jere R, Susan W Parker, and Petra Todd. 2020. "Impacts of Prospera on Enrollment, School Trajectories and Learning." *Working Paper*.

Behrman, Jere R, Piyali Sengupta, and Petra Todd. 2005. "Progressing Through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico." *Economic Development and Cultural Change* 54 (1): 237–275.

Bourguignon, François, Francisco HG Ferreira, and Phillippe G Leite. 2003. "Conditional Cash Transfers, Schooling, and Child labor: Micro-Simulating Brazil's Bolsa Escola Program." *The World Bank Economic Review* 17 (2): 229–254.

Buscha, Franz, Arnaud Maurel, Lionel Page, and Stefan Speckesser. 2012. "The Effect of Employment While in High School on Educational Attainment: A Conditional Difference-in-differences Approach." *Oxford Bulletin of Economics and Statistics* 74 (3): 380–396.

Dubin, Jeffrey A, and Daniel L McFadden. 1984. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica: Journal of the Econometric Society*, pp. 345–362.

Dubois, Pierre, Alain De Janvry, and Elisabeth Sadoulet. 2012. "Effects on School Enrollment and Performance of a Conditional Cash Transfer Program in Mexico." *Journal of Labor Economics* 30 (3): 555–589.

Eckstein, Zvi, and Kenneth I Wolpin. 1999. "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities." *Econometrica* 67 (6): 1295–1339.

Epple, Dennis, Akshaya Jha, and Holger Sieg. 2018. "The superintendent's Dilemma: Managing School District Capacity as Parents Vote with their Feet." *Quantitative Economics* 9 (1): 483–520.

Ferreyra, Maria Marta. 2007. "Estimating the Effects of Private School Vouchers in Multidistrict Economies." *American Economic Review* 97 (3): 789–817.

Gottfried, Michael 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 (2): 434–465.

Keane, Michael P, Sonya Krutikova, and Timothy Neal. 2018. "The Impact of Child Work on Cognitive Development: Results from Four Low to Middle Income Countries." *IFS Working Papers*.

Le Barbanchon, Thomas, Diego Ubfal, and Federico Araya. 2020. "The Effects of Working while in School: Evidence from Uruguayan Lotteries." *Working Paper*.

Leite, Phillippe, Ambar Narayan, and Emmanuel Skoufias. 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, no. 1.

Neilson, Christopher. 2014. "Targeted Vouchers, Competition Among Schools, and the Academic Achievement of Poor Students." *Working Paper*, pp. 1–62.

Neilson, Christopher, Claudia Allende, and Francisco Gallego. 2019. "Approximating the Equilibrium Effects of Informed School Choice." *Working Paper*.

Parker, Susan W, and Petra E Todd. 2017. "Conditional Cash Transfers: The Case of Progresa/Oportunidades." *Journal of Economic Literature* 55 (3): 866–915.

Schultz, T Paul. 2004. "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program." *Journal of Development Economics* 74 (1): 199–250.

Stinebrickner, Ralph, and Todd R Stinebrickner. 2003. "Working During School and Academic Performance." *Journal of Labor Economics* 21 (2): 473–491.

———. 2008. "The Causal Effect of Studying on Academic Performance." *The BE Journal of Economic Analysis & Policy* 8, no. 1.

Todd, P., and K. Wolpin. 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, Petra, and Kenneth I Wolpin. 2018. "Accounting for Mathematics Performance of High School Students in Mexico: Estimating a Coordination Game in the Classroom." *Journal of Political Economy* 126 (6): 2608–2650.

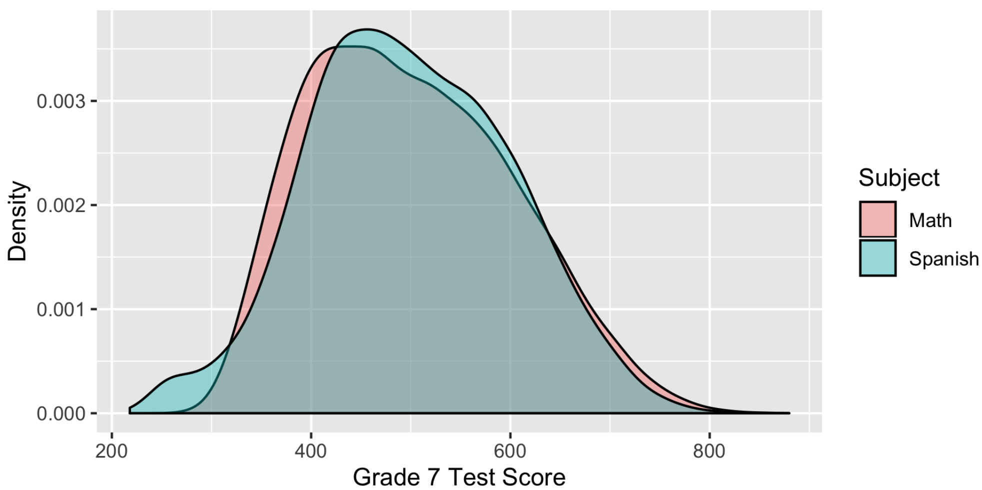
Yap, Yoon-Tien, Guilherme Sedlacek, and Peter F Orazem. 2009. "Limiting Child Labor through Behavior-based Income Transfers: An Experimental Evaluation of the PETI Program in Rural Brazil." In *Child Labor and Education in Latin America*, 147–165. Springer.

Yuan, Ke-Hai, Ying Cheng, and Jeff Patton. 2014. "Information Matrices and Standard Errors for MLEs of Item Parameters in IRT." *Psychometrika* 79 (2): 232–254.

Appendices

A. Histograms of Raw Test Scores

Figure A-1
Grade 7 Test Score Distribution



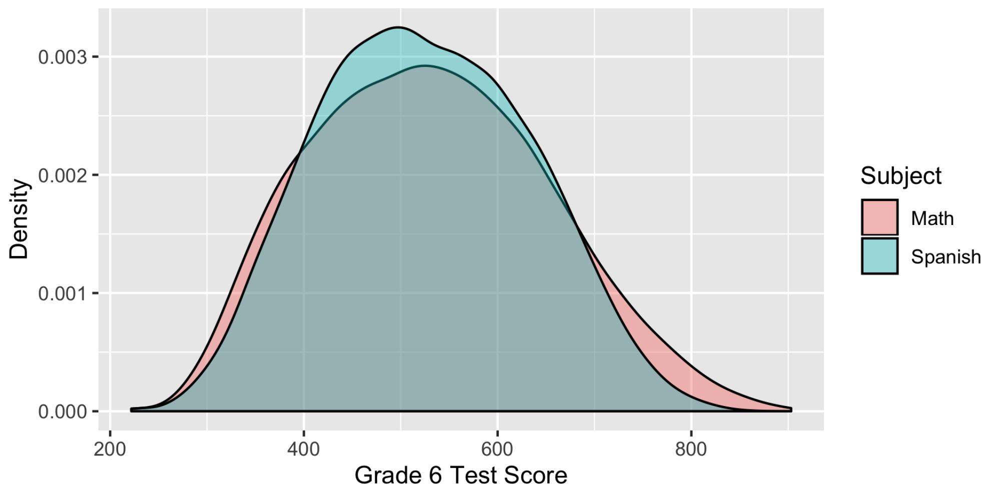
B. Factor Analysis for Effort Questions

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,

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

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

Figure A-2
Grade 6 Test Score Distribution



variables have an underlying joint continuous distribution. The polychoric correlation matrix for my five measures of effort is calculated to be:

Table B-1
Polychoric Correlation Matrix for Effort Variables

	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,

Table B-2
Loading Factors for the Effort Variables

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 2 shows a histogram of the final effort measures.

C. Estimation Strategy Details

1. Guess parameters. There are 51 parameters in this version of the model:
    - 15 coefficients for each of the achievement value added equations
    - 3 parameters in the variance-covariance matrix for the value added equations

- 17 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 2.
   - Compute expected math and Spanish scores using effort and Equation 1 (without the error terms since it is an expectation).
   - 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.

D. Creating Estimation Sample

For students to be in my main estimation sample, I require data on their Grade 6 and 7 ENLACE tests, as well as survey responses from the student and their parent in Grade 7. I am using the survey responses in Grade 7, since I need to know if the students are working or not in that year.

Unfortunately, this sample excludes any students who dropped out between Grade 6 and Grade 7, and I want to model this behavior as well. To incorporate these students into the sample, I randomly select students who wrote the ENLACE tests in Grade 6, have student and parent surveys from Grade 6, and dropped out after Grade 6 and include them in my estimation sample. The number of students I include is chosen so that the dropout ratio is the same as in the full dataset. In doing this, I am assuming that some of the background information from the survey, such as parental education, are constant over these two years.

E. Calculating the Standard Errors

Standard errors are calculated using a sandwich-type covariance matrix (Yuan, Cheng, and Patton 2014). Define the log likelihood for student i given parameters Ω as $L_i(\Omega)$. As detailed in the estimation section, I am able to calculate such probabilities using the data and parameters. To estimate the covariance matrix with a sample of n students, I use the following formula:

$$\hat{\text{Cov}} = \frac{\hat{A}^{-1}\hat{B}\hat{A}^{-1}}{n} \tag{E-1}$$

where

$$\hat{A} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 L_i(\hat{\Omega})}{\partial\hat{\Omega}\partial\hat{\Omega}'}$$
$$\hat{B} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial L_i(\hat{\Omega})}{\partial\hat{\Omega}}\right]\left[\frac{\partial L_i(\hat{\Omega})}{\partial\hat{\Omega}}\right]'$$

The matrix $\hat{A}$ is an estimation for the Hessian, and the matrix $\hat{B}$ is an estimation of the outer product of the gradient. I calculate the gradient and the Hessian numerically in R, using the functions *grad()* and *hessian()*. To get the final standard errors, I take the square root of the diagonal elements of the covariance matrix.

F. 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 mother in number of years MX2010A_EDATTAIN_MOM)
- Educational attainment level of father 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)

The following table lists the data cleaning steps, and the associated sample size.

Sample Size	% Decrease	Notes
11,938,402	-	Raw data downloaded from IPUMs - 2010 census.
11,931,302	0.06%	Drop observations with no recorded wage.
1,812,587	84.8%	Drop all observations with age less than 12 or more than 18.
1,808,019	0.00%	Drop all observations with school attendance not recorded as enrolled or dropped out.
1,800,341	0.01%	Drop those who have graduated school (years of school greater than 13).
1,779,884	0.01%	Drop those with monthly income reported in the top 1%.
1,768,470	0.01%	Drop those with income per hour reported in the top 1%.
1,752,385	0.01%	Drop those with weekly hours of work reported in the top 1%.

Some other data cleaning details, which did not limit the sample size:

- Those with missing or unknown values for monthly personal income have their income set to zero. The same thing was done for monthly family income.
- For the children who do not have parental education recorded, a new variable is created to indicate that they are missing this observation. In doing so, they can still be included in the analysis. 12.6% of the data do not have mother's education recorded, and 25.7% are missing father's education.
- A northern dummy variable is created for all states that are in the northern region of Mexico.
- A dummy variable is created to indicate if the child is working or not. A child is considered to be working if they work more than 5 hours per week, if they report an income greater than zero, and if they report having a job or being employed.
- A net family income variable is created. This is done by subtracting the child's income from the family income.

To account for non-random selection into working, a two-step Heckman selection model is estimated. The first step is a probit model on the probability of working. The full dataset is used to estimate this probit, and it is estimated separately by gender. For boys, there are 870,158 observations and of these 90,549 (10.4%) are recorded as working. For girls, there are 882,227 observations and of these 39,255 (4.4%) are recorded as working. The wage regressors include: age, school attendance, educational attainment, parental educational attainment, parents educational attainment is missing, urban-rural dummies, north-south dummies, and municipality dummies. Interaction terms of all of the above variables are incorporated as well. 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 monthly wages as the independent variable, and the regressors listed in the probit (without the exclusion restriction variables). The model includes municipality fixed effects, which allow for 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 F-1.

Table F-1
Wage Regression Results: Monthly Wages

Dependent Variable:	Monthly Income			
Model:	Boys		Girls	
Age	109.6***	(8.997)	47.31***	(15.86)
School enrollment	-1,162.8***	(127.6)	-977.6***	(202.9)
Mom education missing	-91.84	(165.3)	62.65	(261.3)
Dad education missing	-153.7	(129.0)	-69.75	(205.1)
Educational attainment	59.61***	(20.95)	-79.05**	(32.84)
Mom's education	-66.65***	(18.15)	-64.29**	(28.52)
Dad's education	-44.22**	(19.17)	-24.05	(30.70)
Urban	-1,061.0***	(201.7)	-578.5*	(319.1)
Mills ratio	9.203**	(3.820)	13.34**	(6.062)
Age × School enrollment	21.29***	(8.142)	19.05	(12.33)
Age × Mom education missing	10.09	(9.923)	-8.911	(16.01)
Age × Dad education missing	10.29	(7.858)	3.770	(12.55)
Age × Educational attainment	-2.779**	(1.263)	5.738***	(1.969)
Age × Mom's education	3.884***	(1.108)	4.062**	(1.737)
Age × Dad's education	2.918**	(1.163)	1.644	(1.856)
Age × Urban	70.66***	(12.40)	35.20*	(19.66)
School enrollment × Urban	144.7	(154.3)	174.3	(254.0)
Mom education missing × Urban	-181.3	(225.6)	-648.6**	(330.7)
Dad education missing × Urban	113.5	(185.3)	323.4	(276.2)
Educational attainment × Urban	7.696	(28.21)	46.49	(42.16)
Mom's education × Urban	14.94	(21.88)	1.520	(34.44)
Dad's education × Urban	40.96*	(24.35)	-4.873	(36.61)
Age × North	-10.35	(24.85)	27.60	(45.36)
School enrollment × North	-129.3	(287.1)	-1,426.7***	(378.3)
Mom education missing × North	166.8	(414.4)	798.2	(650.2)
Dad education missing × North	-433.5	(374.8)	-638.5	(581.1)
Educational attainment × North	-54.12	(52.29)	18.25	(96.79)
Mom's education × North	34.46	(35.60)	23.22	(60.62)
Dad education missing × North	-53.63	(38.06)	-39.49	(61.69)
Urban × North	0.4240	(35.70)	-39.07	(58.94)
Age × School enrollment × Urban	-6.560	(9.815)	-11.10	(15.53)
Age × Mom education missing × Urban	13.40	(13.57)	45.45**	(20.07)
Age × Dad education missing × Urban	-6.835	(11.27)	-17.88	(16.81)
Age × Educational attainment × Urban	-0.9107	(1.687)	-2.880	(2.515)
Age × Mom's education × Urban	-0.7190	(1.338)	-0.5548	(2.091)
Age × Dad education missing × Urban	-2.869*	(1.476)	0.0820	(2.206)
Age × School enrollment × North	12.38	(17.92)	88.89***	(23.46)
Age × Mom education missing × North	-12.09	(24.83)	-47.28	(38.05)
Age × Dad education missing × North	25.97	(22.29)	34.61	(34.39)
Age × Educational attainment × North	3.230	(3.077)	-2.110	(5.628)
Age × Mom's education × North	-2.053	(2.163)	-1.748	(3.574)
Age × Dad education missing × North	3.247	(2.293)	2.124	(3.637)
Fixed-effects				
Municipalities	Yes		Yes	
Fit statistics				
Observations	89,221		38,769	
R^2	0.30957		0.39039	
Within R^2	0.18828		0.16758	

Clustered (Municipalities) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*