Constraints on Children's Education: Enrollment, Effort, and Work Decisions in Mexico*

Gabrielle Vasey†

When children work, their education competes for their time and effort. This paper develops and estimates a structural model in which Grade 6 students in Mexico choose among school–work alternatives, with study effort determined endogenously. Choices depend on local school options, travel distance, and labor market wages. Achievement is modeled using a value-added specification with effort as an input. The model is estimated using test scores, survey data from students, parents, and principals, school locations, and census-based wage data. Counterfactual simulations show that enforcing child labor bans would reduce dropout between Grades 6 and 7 by only 6.6 percent, with modest gains in effort and test scores. Policies that reduce travel costs or improve perceptions of rural telesecundarias achieve larger enrollment gains at potentially lower cost. The results highlight the importance of incorporating both schooling and labor choices when evaluating education policies in contexts where child labor is prevalent.

JEL Codes: I21, I25, J22, O15.

Keywords: Education Policy, School Choice, Child Labor, Study Effort, Structural Model, Conditional Cash Transfers

*I would like to thank Jere Behrman, Jorgen Hansen, Susan Parker, Holger Sieg, Petra Todd, and Arthur van Benthem 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. I thank Hector Robles Vaquez for preparing databases and for assistance working with these data. I am 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

Around the world, students often face significant constraints when deciding whether to continue their education. These include the opportunity cost of schooling, limited access to nearby schools, and the challenge of balancing study time with other responsibilities, such as work or household tasks. Even in countries where school enrollment is legally mandated and child labor is restricted, enforcement is frequently weak, and children must weigh the short-term benefits of contributing to household income against the long-term returns to education. For many students, the decision is not simply whether to enroll, but also how much effort to invest once in school and whether to combine school with work.

This paper studies how students make enrollment and effort decisions when transitioning to middle school in a setting where they face multiple constraints. In Mexico, students must choose a new school after completing primary school (Grade 6), creating a natural context for studying school choice under limited access, work obligations, and effort costs. Although education through Grade 9 is compulsory and labor for children under 14 is prohibited by law, enforcement is uneven. In the 2010 Census, 7.9% of 12- and 13-year-olds were not enrolled in school. Among those who were enrolled, over a quarter reported working at least one day per week. These figures suggest that formal rules alone are not sufficient to keep students in school, and that decisions about enrollment and effort are shaped by both policy and tradeoffs.

To analyze these decisions, I develop and estimate a model in which students jointly choose whether to enroll in school, whether to work part time, and how much effort to exert if enrolled. These decisions are made in the context of a rich and heterogeneous choice environment. After Grade 6, each student faces a set of available middle schools, defined by distance, school type, and perceived quality. Schools are modeled as differentiated options, and students consider both academic and non-academic factors in their decisions. Effort is costly, and its marginal cost depends on individual characteristics and labor status. Students also receive wage offers, which vary by location and demographic profile, capturing the opportunity cost of time spent in school.

A key feature of the model is that it allows students to combine school and work, rather than treating these as mutually exclusive choices. This more realistic structure captures the full range of time allocation decisions that students face and highlights the potential tradeoffs between earning income and investing in academic success. While a few prior studies consider joint school–work decisions, they typically do not model academic outcomes alongside them (Bourguignon, Ferreira, and Leite 2003; Leite, Narayan, and Skoufias 2015).

To estimate the model, I combine administrative data on nationwide standardized tests in math and Spanish with detailed survey data from students, parents, and principals, as well as geocoded school locations and census data on local labor markets. Together, these sources create an unusually rich dataset that allows me to observe enrollment decisions, academic performance, reported study effort, and work behavior for a large national sample. The data also indicate which students receive the conditional cash transfer Prospera. These linked sources have recently been used to evaluate educational interventions in Mexico,[1] and they provide the foundation for estimating a model that can simulate the effects of alternative policies.

The model is a discrete-continuous choice framework in which students select from discrete alternatives, school type and work status, while also choosing a continuous level of study effort. It builds on a random utility structure (Dubin and McFadden 1984), with effort determined as the optimal response within each school–work alternative. Academic achievement is modeled using a value-added production function that depends on both effort and school characteristics.[2] I estimate the model by Maximum Likelihood, exploiting the fact that the likelihood can be decomposed into three conditional components, each of which has a closed-form solution.

Identification in the model is supported by rich geographic variation in school access and labor market conditions. To recover the value students place on school characteristics, I use variation in the distance to nearby schools and differences in school type and quality within students' choice sets. Local wage offers, which vary by demographic profile and municipality, identify the opportunity cost of schooling. Effort plays a central role in the model, and parameters governing its cost and productivity are identified using self-reported measures of study behavior, which vary systematically across students and contexts.

The results show that constraints, particularly travel distance and work opportunities, play a substantial role in shaping enrollment and effort decisions. Students are sensitive to the distance required to attend middle school, and they place significantly lower value on distance education schools (Telesecundarias) compared to General and Technical schools. While students value higher achievement, this preference does not vary meaningfully by parental education or transfer status. Effort is costly, especially when students

[1] See Acevedo, Ortega, and Székely 2019, Behrman, Parker, and Todd 2020, Borghesan and Vasey 2024, Behrman et al. 2025.

[2] There is a large literature both developing and estimating value added production functions in educational settings. Some examples include Boardman and Murnane 1979, Hanushek 1979, Todd and Wolpin 2003, and Andrabi et al. 2011.

are also working, but the cost is lower for female students and those with stronger prior academic performance. Effort is estimated to be a key input into achievement, confirming its importance as a mechanism.

I use the estimated model to simulate how alternative policies that relax access and time constraints would affect enrollment, achievement, and work behavior. These counterfactuals include adjusting conditional cash transfers, reducing travel distance to school, and shifting perceptions about rural school types. I also consider the effect of strictly enforcing child labor laws, though this policy proves less effective than those that address the underlying barriers to enrollment. The results highlight that targeted transfers and improved school access through transportation or better information can lead to meaningful increases in enrollment and effort, with relatively modest costs.

This paper contributes to four strands of research. First, it builds on work studying school choice in settings where students face geographic and institutional constraints, and where schools are treated as differentiated options (e.g., Ferreyra 2007; Epple, Jha, and Sieg 2018; Neilson 2014; Bau 2019; Neilson, Allende, and Gallego 2019). Second, it contributes to the literature on education production by explicitly modeling student effort as a key endogenous input (Todd and Wolpin 2018; Stinebrickner and Stinebrickner 2008). Third, it relates to research on the tradeoffs between education and child labor, particularly in the context of conditional cash transfers in low- and middle-income countries (e.g., Attanasio, Meghir, and Santiago 2011; Todd and Wolpin 2006; Parker and Todd 2017; Alcaraz, Chiquiar, and Salcedo 2012).[3] Finally, it connects to a smaller but important literature on child labor itself, particularly in settings where work is informal, unpaid, or within the household (Bourguignon, Ferreira, and Leite 2003; Leite, Narayan, and Skoufias 2015; Keane, Krutikova, and Neal 2018).[4]

This paper is closely related to prior structural work on schooling and labor decisions in Mexico, notably Todd and Wolpin (2006), Attanasio, Meghir, and Santiago (2011), and Behrman et al. (2025), which develop dynamic models of enrollment, work, and achievement to evaluate the impacts of conditional cash transfer programs such as Progresa/Prospera. Among these, Behrman et al. (2025) is the most recent and methodologically closest, modeling Grades 4–9 and allowing students to work while enrolled. In contrast, the present paper narrows the focus to a single, high-stakes transition, entry into

[3]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).

[4]Other work focusing on child labor in Mexico includes Tagliati 2019, Cummings 2016, and Kozhaya and Flores 2022, who examine trends and interventions targeting labor market participation among school-aged children. While these studies provide useful context, they do not model school effort or academic outcomes directly.

Grade 7, where dropout risk sharply increases. I model this margin in greater detail by jointly incorporating school choice, labor participation, and endogenous academic effort. The use of linked administrative and survey data with direct measures of effort allows for identification of behavioral responses that are not separately captured in the earlier models. This focus on a critical bottleneck, combined with the explicit effort margin, yields policy counterfactuals that differ in magnitude and ranking from those in Behrman et al. (2025).

Within this landscape, the paper makes several contributions. Empirically, it studies a setting where school transitions are mandatory and school access is highly variable, allowing for a clear test of how students make enrollment and effort decisions under constraint. Methodologically, it develops a structural model that jointly incorporates school choice, labor participation, achievement production function, and endogenous effort, which is a combination rarely seen in prior work. The model is estimated using rich linked administrative and survey data from Mexico, capturing academic outcomes, study behavior, and local labor markets. Finally, the paper uses the estimated model to simulate a range of policy scenarios that address different constraints, including school access, opportunity costs, and perceptions of school quality, and examines their effects on enrollment, achievement, and work decisions.

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

II. Data and Setting

This section describes the data sources, sample construction, and key variables used to estimate the model. To capture the joint decisions students make about enrollment, work, and effort, the dataset must include information on school access, academic achievement, labor participation, and time use. The data I use meet these requirements and allow for estimation of the model's key behavioral and structural parameters.

A. Data Sources

To carry out this research, I combine four main data sources: standardized test scores, a nationally representative education survey, census-based labor market data, and administrative records of school locations and characteristics.

The base of the dataset comes from 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. In addition to the tests, 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, Behrman et al. 2025).

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. I also merge information on schools in Mexico, including their exact geographic coordinates.

Finally, the analysis requires data on wages, which are not recorded in the previously mentioned sources. The 2010 Census is used to access information on children between the ages of 12 and 18, including 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.

Together, these sources yield a nationally representative sample of students across Mexico. For each student, I observe standardized test scores, school identifiers (with associated school information), individual and household demographics (including conditional cash transfer status), and the average municipal wage conditional on age, gender, family background and school attendance.

B. Estimation Sample

This analysis focuses on students in Mexico who were enrolled in Grade 6 in 2008 and could potentially transition to Grade 7 in 2009.[5] The sample can be divided into two groups: those who enrolled in school in Grade 7, and those who dropped out of school

[5]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.

after Grade 6.[6] Among the 229,199 students enrolled in Grade 6 in 2008 with matched parent and student survey data, 17,195 (7.5%) do not appear in either the ENLACE or enrollment registry data over the next four years. I classify these students as having dropped out.

In Grade 7 in 2009, there are 107,898 students for whom we have survey answers from themselves and their parents. 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.[7] This grade is particularly relevant because it marks the first year of secondary school, when students must select a new school and face new constraints related to access and opportunity cost. 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 balanced by gender, as 49.9% of the students are female. 26.1% of students are beneficiaries of the conditional cash transfer Prospera.

C. Key Variables

This section describes the key variables used to estimate the structural model of school, work, and effort decisions. These variables capture both the outcomes of interest (achievement, effort, work status) and the key features of the decision-making environment (school characteristics, access, and local wages).

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.[8] The same transformation was used in subsequent years.[9] All students write the test in Grade 6 and Grade 7, so it is possible to control for lagged test scores.

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

[7] Sample size also changes from year to year.

[8] The distributions of the test scores are approximately normal, as shown in Appendix A.

[9] Since the same transformation is used each year, subsequent years do not necessarily have a mean of 500 or standard deviation of 100. Increases in the mean year upon year are interpreted as increases in learning.

C..2 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.

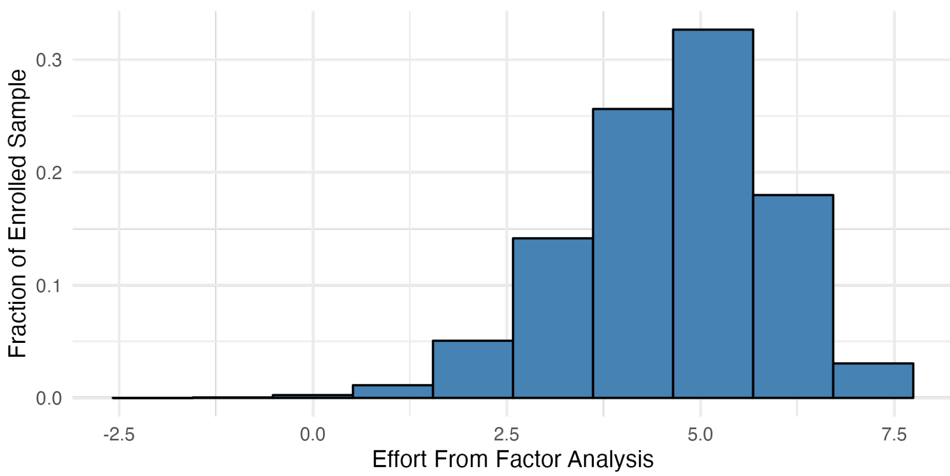
The first measure, the number of hours studied per day, is cardinal. The other four measures are ordinal variables, answered on a Likert scale. To combine them into one value, I use factor analysis. This analysis is done outside of the model estimation, and uses polychoric correlations to take into account 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 which combines the information from the student's responses to the five effort questions. Figure 1 presents a histogram of the new continuous effort variable. The effort values are almost all positive and the distribution appears to be approximately normal.[10] Estimation details and results are in Appendix B.

C..3 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 2 shows the responses, divided by gender. Boys work more than girls, and the majority of students

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

Figure 1
Latent Effort Variable Distribution



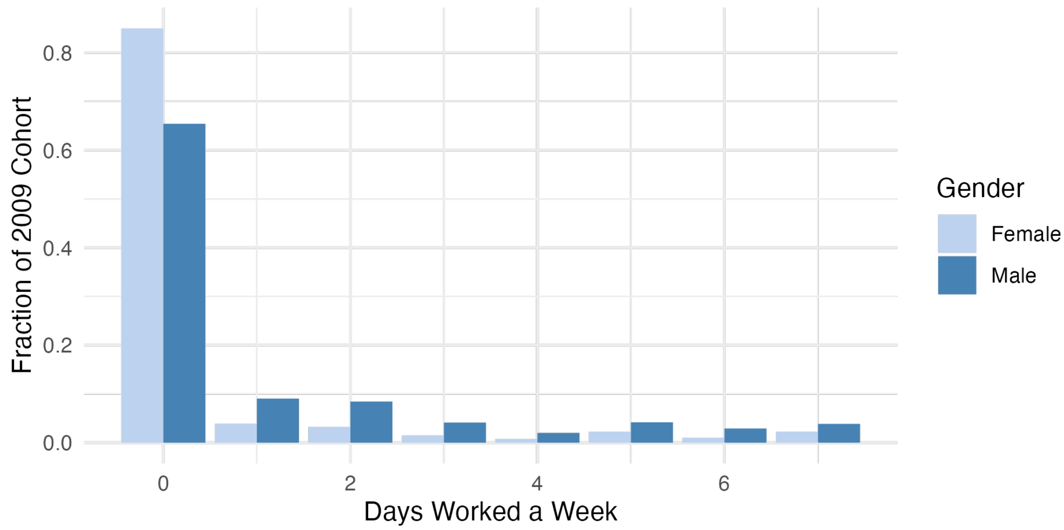
Note: The distribution of latent effort values in the estimation sample. This variable is combines information from five questions related to study effort using factor analysis.

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

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, the labor question does not explicitly distinguish between paid and unpaid work, which introduces some ambiguity in how students interpret and respond to it. There is an additional question inquiring about the reasons for working, and 59% of students reported working for their family. While the high rate of family-based work suggests that the survey captures at least some forms of unpaid labor, underreporting is likely, especially among younger students for whom labor is legally restricted. As a result, the measure of work used in the analysis may understate the prevalence of unpaid or informal labor.

Figure 2
Days Worked per Week



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

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.

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 1 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 1717 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 1
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)

Note: 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.

My approach builds on a wage imputation model used in Behrman et al. 2025, who apply a similar strategy to estimate child wages in the context of school enrollment and achievement in Mexico. Their method separately models monthly hours worked and hourly wages using two regressions. In contrast, I directly estimate monthly wages in a single step, which may reduce measurement error in the imputed values, particularly given the monthly wage format of the census data. While the overall strategy is similar, my implementation is tailored to the needs of the structural model developed in this paper.

One limitation of this approach is that since I am estimating monthly wages, I am not incorporating unpaid work, so children who report working for their families without pay are not classified as "working" in the wage regressions. This raises the concern that the imputed wages may understate opportunity costs in areas where informal or unpaid labor is prevalent. In the structural model, however, I do not exclude students who report working for their families. Instead, I assume that any student who engages in work, whether paid or unpaid, is doing so because their time is worth at least the prevailing market wage in their area. That is, if a student works at home rather than in the labor market, their contribution is assumed to be equally valuable to the household. This assumption allows the imputed wage to serve as the shadow price of a student's time, regardless of the formality or compensation of their work. While this may not capture all variation in informal labor conditions, it provides a consistent and tractable measure of opportunity cost across the estimation sample.

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. Since school tuition was not available, students attending private schools are not included in the estimation of the model.

Table 2 contains summary statistics for the three 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.

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

Table 2
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

Note: 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).

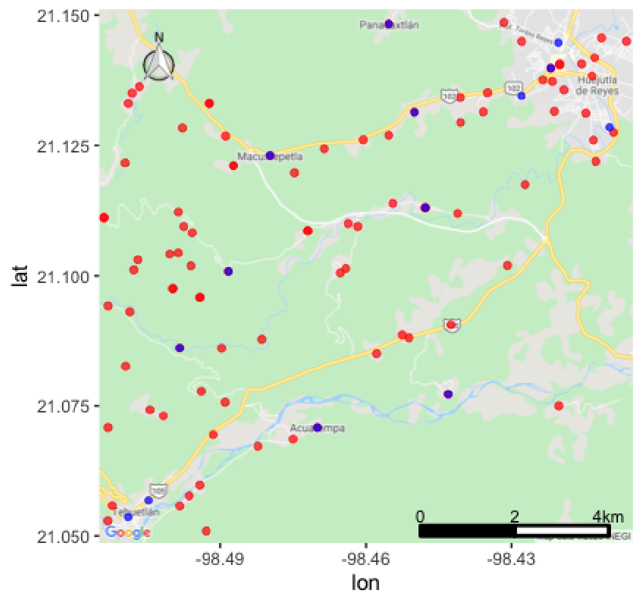
are much more sparse than primary schools, especially in rural regions of Mexico. Figure 3 presents an example, and shows the geographic distribution of primary and middle schools in a region in Mexico. Although there is a small city in the top right corner, the remainder of area covered by the map is 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 assume 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.[11]

For the estimation, I define each student's choice set of middle schools based on the location of their primary school. I draw a circle around each primary school and include all middle schools within that circle in the student's choice set, as illustrated in Figure 4. Rather than using a fixed radius, which would ignore variation in geography and school density, I calculate a custom radius for each primary school using the observed choices of students in the estimation sample. Specifically, the radius is equal to the maximum distance that I observe any student from that primary school traveling to attend middle

[11] It is also possible to calculate distance using roads and paths on Google Maps, but this does not capture many of the rural pathways.

Figure 3
Map with Schools in Example Neighborhood



Note: 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.

school.[12]

D. Data Patterns

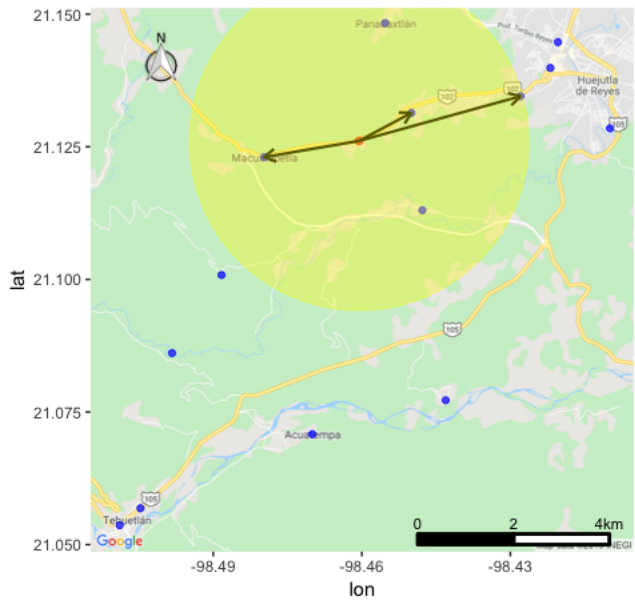
This section presents key empirical patterns that motivate the structural model and highlight the constraints students face when making enrollment and effort decisions. I document relationships between work, effort, achievement, wages, and access to schools. These patterns guide the modeling choices in the next section and serve as important benchmarks for validating the fit of the model in the Results section.

D..1 Working and Achievement

One of the main concerns about child labor is that it may reduce students' academic achievement. Table 3 presents a simple regression showing a significant negative correla-

[12] To prevent the radius from being overly sensitive to outliers or residential moves, I impose a minimum radius of 3 km and a maximum of 20 km. While this approach reflects realized school choices, alternative strategies such as using historical choice data could reduce potential endogeneity in how the choice set is defined.

Figure 4
Map Showing Choice Set in Example Neighborhood



Note: 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.

tion between working and Grade 7 test scores. The variable "Working" is a dummy equal to 1 if the student reports working at least a certain number of days per week: column (1) uses 1 day, column (2) uses 2 days, and so on. While this analysis is descriptive and not intended to be causal, the regressions include controls for lagged test scores, student gender and age, parental education levels, and whether the student is a beneficiary of the conditional cash transfer.

The estimated relationship grows in magnitude as the number of days worked increases. Students who report working at least three days per week score 5.8 percent of a standard deviation lower than their peers. These patterns highlight the importance of accounting for the time cost of work in the structural model.

D..2 Working and Study Effort

To better understand the relationship between working and achievement, it is helpful to examine the potential mechanisms. One likely channel is time and energy: students who work may have less capacity to study. I capture this through a composite measure

Table 3
Days Worked per Week and Test Scores

	Dependent variable:			
	Grade 7 Test Scores (Standardized)			
	(> 0 Day)	(> 1 Days)	(> 2 Days)	(> 3 Days)
	(1)	(2)	(3)	(4)
Working	-0.026^{***}	-0.038^{***}	-0.058^{***}	-0.082^{***}
	(0.007)	(0.008)	(0.009)	(0.010)
Lagged Test Scores	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	64,215	64,215	64,215	64,215
R^2	0.506	0.506	0.507	0.507

Note: 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 standardized sum of each student's Grade 7 math and Spanish test scores. Controls include female, age, parent education, and CCT status.
** Significant at 5% level, *** Significant at 1% level.

of study effort, described in detail in Section C..2.

Table 4 shows a negative correlation between working at least one day per week and the effort variable. The magnitude of this relationship declines slightly as controls are added, but remains statistically significant. In the specification with the full set of controls, students who report working have effort scores that are 5.7 percent of a standard deviation lower than those who do not. This pattern supports the hypothesis that labor may crowd out academic effort, which in turn affects achievement.

Table 4
Working and Study Effort

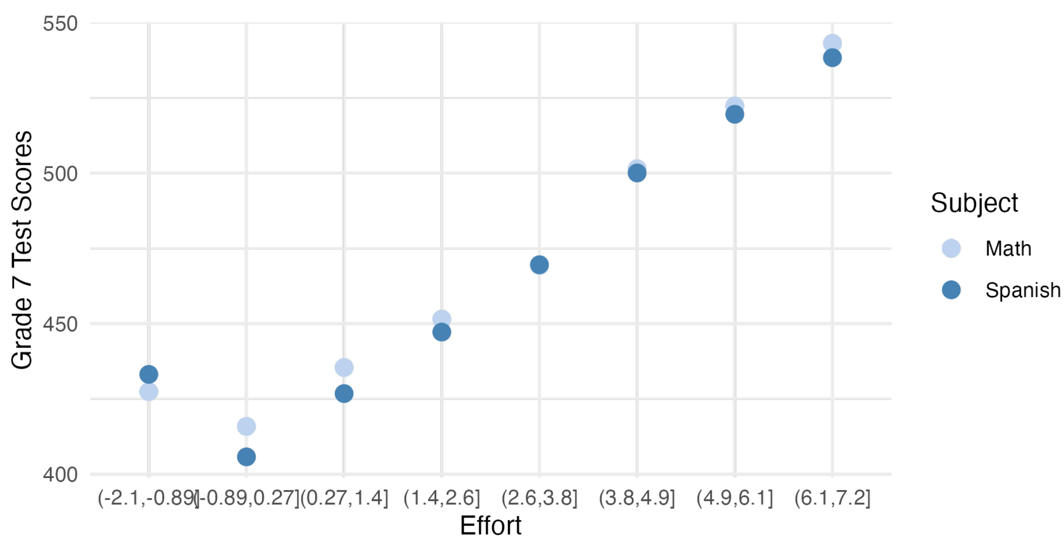
	Dependent variable:		
	Effort (Standardized)		
	(1)	(2)	(3)
Working	-0.160^{***}	-0.069^{***}	-0.057^{***}
	(0.009)	(0.009)	(0.009)
Lagged Test Scores	No	Yes	Yes
Controls	No	No	Yes
Observations	64,215	64,215	64,215
R^2	0.005	0.065	0.077

Note: Correlation between working 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 standardized continuous variable created with factor analysis. Controls include female, age, parent education, and CCT status.
** Significant at 5% level, *** Significant at 1% level.

D..3 Study Effort and Achievement

For study effort to be a plausible mechanism between working and achievement, it must itself be predictive of test performance. 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 5 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.

Figure 5
Study Effort and Test Scores



Note: 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.

D..4 School Accessibility and Enrollment

When deciding whether to enroll in school or not, students face both opportunity costs and physical constraints. One key constraint is the distance to the nearest middle school. The farther away a middle school is from their primary school, the higher the traveling cost.[13]

Figure 6 shows a clear negative relationship between school accessibility and enrollment. For each primary school, I calculate the distance to the nearest middle school and the proportion of students who did not enroll in Grade 7. Students whose primary schools are farther from any middle school are significantly more likely to drop out. This geographic variation in access is an important feature of the environment and is explicitly incorporated into the school choice model.

[13]Students may still choose a school that is further away because of its type or quality.

Table 5
Study Effort and Test Scores

	Dependent variable:		
	Grade 7 Test Scores (Standardized)		
	(1)	(2)	(3)
Study Effort	0.306***	0.135***	0.130***
	(0.004)	(0.003)	(0.003)
Lagged Test Scores	No	Yes	Yes
Controls	No	No	Yes
Observations	64,215	64,215	64,215
R^2	0.094	0.519	0.522

Note: Correlation between study effort and test scores in the data. Test scores are the standardized sum of Grade 7 math and Spanish scores. The effort variable is the standardized continuous variable created by factor analysis. Controls include female, age, parent education, and CCT status.
** Significant at 5% level, *** Significant at 1% level.

D..5 Wages and Working

Students' work decisions are influenced not only by their own characteristics but also by the local labor market. Because wage opportunities vary by location, the imputed wage variable introduces geographic variation in the returns to working. Table 6 shows results from a linear probability model where the dependent variable is a student's dropout decision (equal to 1 if they dropped out between grade 6 and grade 7). A one standard deviation increase in the imputed wage is associated with a 1.5 percentage point increase in the probability of dropping out. This effect is substantial given that the national dropout rate at this transition is under 8 percent.[14]

These results highlight the importance of accounting for heterogeneity in wage offers when modeling school enrollment decisions. The model incorporates this heterogeneity by allowing wages to vary with student characteristics and municipality of residence.

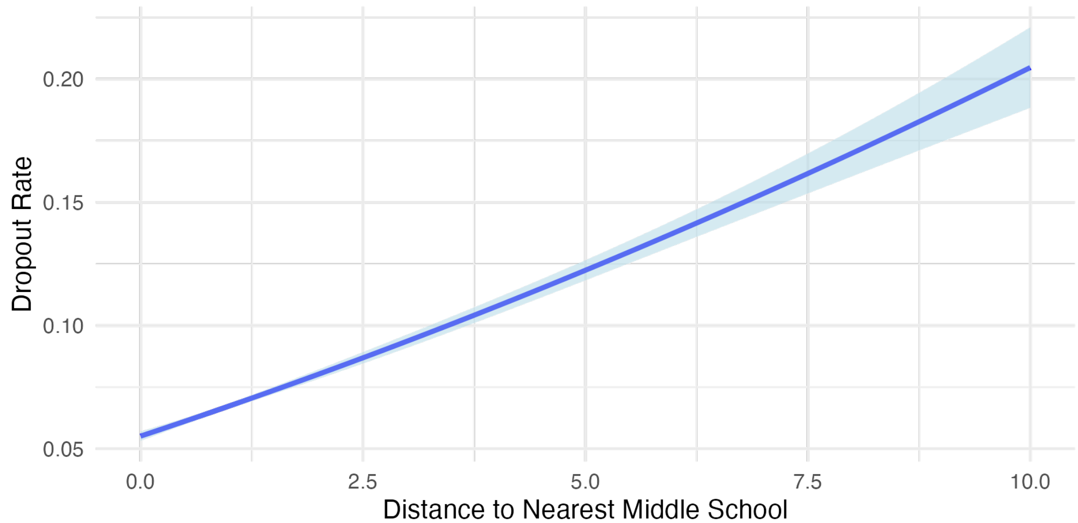
[14] Up until the end of primary school (grade 6), there is little dropout. Grade 7 is the first year of secondary school.

Table 6
Dropout and Wages

	Dependent variable:		
	Dropout		
	(1)	(2)	(3)
Imputed Wage	0.017***	0.015***	0.015***
	(0.001)	(0.001)	(0.001)
Lagged Test Scores	No	Yes	Yes
Controls	No	No	Yes
Observations	69,382	69,382	69,382
R^2	0.004	0.028	0.039
Note:			*p<0.1; **p<0.05; ***p<0.01

Note: Correlation between dropping out and the imputed monthly full time wage. The sample used is the main estimation sample. Wages are normalized to have a standard deviation of one Controls include female, parent education, urban, and CCT status.
** Significant at 5% level, *** Significant at 1% level.

Figure 6
Enrollment Rates and Access to Middle Schools



Note: The relationship between dropout rates and middle school accessibility. For each primary school, two values are calculated: the distance to the nearest middle school (x-axis) and the fraction of students who do not enroll in Grade 7 (y-axis).

III. Model

The model captures the key decisions that students make as they transition from primary school (Grade 6) to middle school (Grade 7). In Grade 6, each student i is enrolled in a primary school P_i, which determines their local schooling environment. Students are defined by a set of characteristics (gender, age, parental education), their lagged test scores, and if they are a conditional cash transfer beneficiary.

Based on the location of their primary school, each student faces a set of accessible middle schools $\mathbb{S}_{P_i}$. Middle schools are modeled as differentiated products, varying by type (General, Technical, or Telesecundaria), geographic distance from the student's primary school, and observable school-level characteristics.

At the time of decision-making, students choose (i) whether to enroll in school, (ii) whether to work part-time if enrolled (or full-time if not), and (iii) how much academic effort to exert if enrolled. These choices are made jointly, conditional on the student's choice set and wage offers. Wages vary by demographic profile, location, and school enrollment status, and reflect the opportunity cost of time.

A central feature of the model is that school, work, and effort decisions are jointly determined and interact. Unlike models that treat work and school as mutually exclusive, this framework allows students to combine part-time work with school enrollment, and accounts for how this tradeoff influences effort and achievement. Effort is costly and heterogeneous, but also directly improves academic outcomes via a production function. The model allows estimation of how constraints such as limited school access or high opportunity costs shape behavior and outcomes through multiple margins.

A. Student Utility

Students in the model are, on average, 12 years old, making it plausible that school decisions are made jointly with their families. I do not explicitly model a household utility function or a family budget constraint; instead, I specify utility at the level of the student. This simplification is consistent with a setting in which family utility is additively separable across members and parents act to maximize the utility of their children. While an explicit household budget constraint would be ideal, the income data in the survey is reported in coarse categorical bins that are not suitable for estimation. To partially account for household resource constraints and differences in parental priorities, I allow preferences over achievement and school characteristics to vary with parental education, which serves as a proxy for both economic and informational advantages. This approach provides flexibility in how families weigh academic benefits and school attributes, even in the absence of direct income measures.

Students and their families are assumed to care about the student's academic achievement, monetary compensation (transfers and wages), the type and proximity of the school, and the cost of exerting academic effort. Effort allowed to be more or less costly for students with lower prior achievement or other time demands, such as part-time work. The utility of student i attending school j and selecting labor option L, combined with their chosen level of effort e, is given by:

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

The monetary compensation component includes two elements: the conditional cash transfer CCT_i which student i receives if they are a Prospera beneficiary, and the part-time wage w_i^{PT}, which they receive if they choose to work while enrolled in school. The coefficient on the monetary component is normalized to one, so that the units of all other coefficients in the utility function can be interpreted in monetary units (pesos).

The cost of accessing a school is modeled as a function of the straight-line distance between student i's primary school P_i and middle school j, denoted $d_{P_i j}$, and enters utility quadratically to allow for diminishing sensitivity.

Expected academic achievement, $\hat{A}_{ij}^7(e_{ijL})$, depends on the student's effort e_{ijL}, their characteristics, and the characteristics of school j. The value that families place on achievement is allowed to vary with parental education and CCT status. Specifically, parental education is captured by an indicator $PEduc_i$ equal to one if both parents have completed at least middle school.

A general preference for school enrollment is captured by a fixed benefit α_6, which also varies with parental education through an interaction term. Preferences over school type are captured by a set of indicator variables for each of the three school types (General, Technical, and Telesecundaria), with corresponding coefficients β_k.

Finally, students may experience disutility from combining part-time work and schooling. This is captured by a fixed penalty term α_8 when a student selects part-time work while enrolled.

The utility function includes both a linear and a quadratic term in effort. This functional form allows for flexible responses to incentives and guarantees an interior solution for each student's optimal effort choice. The marginal cost of effort varies across students through a random coefficient $\alpha_{i,9}$, which captures heterogeneity in how burdensome academic effort is perceived to be. This coefficient is modeled as a function of both observed and unobserved student characteristics:

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

where X_i includes student gender, parental education, and lagged test scores, and the unobserved component η_i is normally distributed: $\eta_i \sim \mathcal{N}(0, \sigma_\eta^2)$.

By placing heterogeneity in the cost of effort rather than in the returns to effort, the model better matches what can be credibly identified in the data. I observe student-reported effort and test scores, and I estimate a test score production function that links achievement to effort, student characteristics, and school inputs. However, allowing for student-level heterogeneity in the returns to effort would be difficult to separately identify. Effort is self-reported and endogenously chosen, and test scores reflect both effort and underlying ability, along with other inputs. By contrast, heterogeneity in effort costs is more directly identified through variation in observed effort choices across students facing different school and work environments. Modeling heterogeneity on the cost side therefore allows the framework to capture meaningful behavioral responses such as how wages, distance, or school quality affect effort without requiring strong assumptions about unobserved ability or production function heterogeneity.

If a student selects the outside option, they are assumed to drop out of school after Grade 6 and enter the labor force full time. In this case, their utility consists of the full-time wage offer w_i^{FT} and an idiosyncratic shock ν_{i0}, so that

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

As with all alternatives, the error terms ν_{ijL} are assumed to be independently and identically distributed with a Type I extreme value distribution. This assumption implies that the choice component of the student's problem has a standard multinomial logit form, with observed heterogeneity (e.g., through wages, distances, and preferences) and unobserved heterogeneity entering through the random cost-of-effort function rather than directly in the choice utility.

Each student i faces a choice set $\mathbb{S}_{P_i}$, consisting of all middle schools located within a fixed radius of their primary school, P_i.[15] Schools in the choice set are characterized by their distance from the primary school d_{P_ij}, their school type $Type_j$, and additional school-level variables from the principal survey that I include in the analysis relate to infrastructure and principal education and teacher quality.

[15] A full description of how choice sets are constructed can be found in Section C..6.

B. Wage Offers

Each student receives two wage offers: one for full-time work and one for part-time work. Accepting the full-time wage implies that the student does not enroll in school, while part-time work can be combined with school attendance. Students may also choose to focus solely on school, in which case they will not receive any wages.

Wage offers are not observed directly in the survey data and are therefore imputed using Mexican Census data, as described in Section C..4. The imputation model allows wages to vary by student age, gender, school enrollment status, parental education, and geographic location (urban/rural and municipality). These imputed wages serve as the opportunity cost of schooling and influence student decisions through both the enrollment and effort margins.

C. Expected Test Scores

For students who choose to enroll in school, their Grade 7 test score is generated by a value-added production function. The student inputs to the production function include lagged test scores, student characteristics (age and gender) and their effort choice. School inputs, Z_j, include the type of school and an index variable related to school resources. The school resources variable is comprised of three inputs: principal education, if the school has internet, 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. The error term ξ_{ij} is not observed by students at the time of their decisions, so they base their choices on expected scores. Working does not directly affect achievement, however, working can impact the cost of study effort, thereby indirectly affecting achievement.

D. Maximization Problem

Each student i solves a nested optimization problem. For each feasible school option j and labor option $L \in \{0, PT, FT\}$ in their choice set, the student first chooses the optimal level of effort e^*_{ijL} that maximizes their utility:

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

Solving the first-order condition of the above maximization problem yields a closed-form expression for optimal effort:

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

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 ($\hat{A}^7_{ij}(e^*_{ijL})$).

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

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

IV. Estimation

This section outlines the estimation strategy used to recover the structural parameters of the model. Students make joint decisions about school enrollment, labor participation, and academic effort, and the data include observed outcomes for achievement and effort. Given this structure, the model is estimated by Maximum Likelihood. The likelihood function takes advantage of the fact that the joint distribution of choices, test scores, and effort can be decomposed into components with closed-form solutions. This approach allows the model to be estimated efficiently while incorporating the full structure of the utility and achievement functions. The remainder of the section defines the joint probability of observed outcomes and describes how it is used to construct the likelihood.

A. Likelihood Function and Estimation

Model parameters are estimated using Maximum Likelihood. Define

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

as the joint probability of choosing school j, labor option L, achieving Grade 7 test score A_{ij}, and reporting 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} follows a Type I extreme value distribution and ξ_{ij} is normally distributed.

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

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

It is useful to decompose the joint probability into a product of conditional probabilities, since each component has a closed-form expression. This structure greatly reduces the computational burden of estimation, as these components can be evaluated directly rather than through simulation. In particular, the probability of achievement conditional on effort and choices, and the probability of observed effort conditional on choices, both depend on the optimal effort level $\tilde{e}^M_{ijL}$, which is computed from Equation 2 given the choice of j and L, the data (X_i, Z_j), the random coefficient shock (η_i), and the model parameters. For notational clarity, conditioning variables are omitted in probability expressions when they are not relevant to the component being shown. The resulting decomposition is presented in Equation 3.

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

The first term is the probability of observing the Grade 7 test score:

$$
P(A_{ij} \mid j, L, X_i, Z_j, \eta_i)
$$

Given the choice j, L, the data (X_i, Z_j), the parameters, and the shock η_i, the optimal effort e^*_{ijL} is computed from Equation 2. The expected test score then follows from Equation 1. Under the assumption that the achievement disturbance ξ_{ij} is i.i.d. normal, test scores are distributed

$$
A_{ij} \sim \mathcal{N}\Big(\widehat{A}^7_{ij}(e^*_{ijL}),\, \sigma^2_\xi \Big)
$$

so the contribution to the likelihood is the corresponding normal density evaluated at A_{ij}.

The second term is the probability of observing the effort measure in the data, conditional on the optimal effort predicted by the model:

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

Optimal effort e^*_{ijL} is given by Equation 2. The coefficient $\alpha_{i,9}$ in the numerator of that expression is a random coefficient with associated shock $\eta_i \sim \mathcal{N}(0, \sigma^2_\eta)$. As a result, effort values are distributed as

$$
\tilde{e}^M_{ijL} \sim \mathcal{N}\big(e^*_{ijL},\, \sigma^2_{e^*} \big)
$$

where $\sigma^2_{e^*}$ is the variance of the true underlying effort. This normal distribution allows the likelihood contribution for observed effort (from factor analysis) to be evaluated in closed form, eliminating the need for simulation when computing the integral in the likelihood.

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

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

Because the idiosyncratic utility shocks ν_{ijL} are assumed to be i.i.d. Type I extreme value, this conditional choice probability has the multinomial logit form:

$$P(j, L \mid X_i, Z_j, w_{ij}, \eta_i) = \frac{\exp(Utility_{ijL})}{\sum_{k=1}^{J_i} \sum_{h \in \{0, PT, FT\}} \exp(Utility_{ikh})}. \quad \textbf{(??)}$$

The term $Utility_{ijL}$ is a function of the optimal effort e^*_{ijL}, the model parameters, and the data. A scale parameter is included in the probability because the coefficient on the monetary component is normalized to one and the outside option is normalized to the value of the full-time wage rather than zero, allowing the overall scale of the utility distribution to be estimated.

Given a set of parameter values and the data, all three probabilities can be computed for each student, and their product defines the individual likelihood. The likelihood in Equation 3 is then maximized to obtain the parameter estimates. Although the choice probability $P(j, L|X_i, Z_j, w_{ij}, \eta_i)$ has the multinomial logit form due to the Type I extreme value assumption on ν_{ijL}, the model is not a conventional mixed logit. The random coefficient η_i enters only through the cost-of-effort function and affects the likelihood via the effort probability term. I use the assumed normal distribution of η_i to evaluate how closely the optimal effort from the model matches the observed effort in the data; this step eliminates the need for numerical integration over η_i in the estimation procedure.

Standard errors are computed using a sandwich-type covariance matrix, as described in Appendix E.. The next subsection discusses the sources of variation that identify the model parameters.

B. Identification

The model parameters are identified primarily through variation in school choice sets and observable characteristics, both across schools and across students within a school. In total, there are 30 parameters to estimate:

- 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

The 12 parameters $\{\delta_k\}_{k=1}^{11}$ and σ_T in the achievement production function are estimated from a value-added equation. The outcome variable is the sum of each student's Grade 7 math and Spanish test scores. Each student also has lagged test scores in both subjects, along with five other covariates. These covariates vary both across schools and across students within a school.

There are 16 parameters in the utility function. Two of these capture the effect of distance, and are identified by geographic variation in distances across students' choice sets. Each primary school has its own set of available middle schools, and each option is associated with a distance and other characteristics. When a particular school type is located far from a given primary school, it may still offer high quality but is chosen by relatively few students (or not at all), providing information on the disutility of travel distance.

Three utility function parameters capture preferences for school type (General, Technical, and Telesecundaria). Given the large number of schools in the data, I do not estimate a separate intercept for each school; instead, the intercept varies by school type. These coefficients are identified by within–choice set variation: students may choose one type of school over another even when it is farther away or offers lower expected test scores, revealing a preference for that type.

Three utility parameters measure how much students value expected test scores. Identification comes from two sources. First, students with higher prior achievement may derive greater utility from attending school rather than dropping out. Second, because achievement depends on school inputs, some schools within a choice set have higher expected test scores, which can increase their attractiveness. Either pattern in the data helps identify the coefficients on expected test scores.

Six utility parameters are associated with the marginal cost of effort. The parameters related to parental education, gender, and lagged test scores are identified from differences in average effort choices across students with these characteristics.

V. Results

The estimated parameters for the utility function are reported in Table 7, while the parameters for the test score production function are shown in Table 8. All utility function coefficients are expressed in units of 100s of pesos per month. Traveling distance to a middle school is estimated to impose a significant cost, with a positive coefficient on the squared distance term indicating that the marginal cost of each additional kilometer declines as distance increases. Both distance-related coefficients are statistically significant.

Table 7
Coefficient Estimates - Utility Function

Coefficients	Estimates	Std.Error
Distance	-45.249	4.006
Distance Squared	1.842	0.163
School Attendance	31.997	13.87
School 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

The estimated benefit of attending school in the utility function is large and increases further when parents have completed middle school.[16] Among school types, Technical schools are valued more highly than General schools, while telesecundarias are perceived significantly less favorably than either of the other two.

The average expected test score has a large, positive coefficient in the utility function, with little variation by parental education or conditional cash transfer status. Given that

[16]This coefficient is equivalent to the negative of the cost of dropping out of school. In the estimation, it is modeled as the cost of dropping out and included in the outside option.

Table 8
Coefficient Estimates - Achievement Function

Coefficients	Estimates	Std.Error
Intercept	1.68	0.38
Lagged Math	0.27	0.02
Lagged Spanish	0.43	0.02
Female	-0.08	0.02
Age	-0.19	0.01
Technical School	0.01	0.03
Telesecondary School	-0.07	0.07
School Resources	-0.12	0.01
Effort	1.57	0.11
Effort X Technical	-0.00	0.00
Effort X Tele	0.02	0.01
Residual Standard Error	1.25	0.00

the standard deviation of test scores is approximately 1, the estimates imply that students and their families value a one-kilometer reduction in distance to school about as much as an improvement in expected test scores of just over 1.5 standard deviations.

There is a strong disutility associated with working part time. In addition, part-time work increases the marginal cost of effort, making effort more costly. The coefficient on effort squared is negative, as required to ensure a well-defined optimum in the model. Effort is estimated to be less costly for female students and for those with higher lagged test scores.

The coefficient estimates in the achievement production function (Table 8) align with expectations. Lagged test scores have a significant effect, with the coefficient on lagged Spanish exceeding that on lagged math. Female students and older students are estimated to perform worse. The value added of Technical schools is essentially the same as that of General schools, whereas telesecundarias appear to have lower value added, though this estimate is not statistically significant. The school resources coefficient is small and negative,[17] while effort has a large, positive, and precisely estimated effect. Effort is estimated to be slightly more productive in telesecundarias than in General and Technical schools,

[17]Of the three measures that make up the school resources variable, the principal's assessment of how often teachers have low performance is likely the most influential, which would be expected to produce a negative effect.

although the difference is small in magnitude.

The magnitude and patterns of the estimates are broadly consistent with prior evidence on school choice and child labor in low- and middle-income countries. The strong negative effect of distance on school enrollment echoes the findings of Duflo (2001), Muralidharan and Prakash (2017), and Andrabi, Das, and Khwaja (2008), who document substantial distance sensitivity in educational choices. In the Mexican context, Attanasio, Meghir, and Santiago (2011) and Todd and Wolpin (2006) also model joint schooling and child labor decisions, and both find that school access has a strong impact on student enrollment decisions. Their policy simulations, however, suggest that reducing travel costs would have only modest effects on enrollment compared to introducing conditional cash transfers. In contrast, my estimates imply somewhat larger effects of reducing travel costs, potentially due to differences in sample composition, school availability, or the joint estimation of school choice, effort, and achievement in my structural model. Relatedly, Behrman et al. (2025) estimate a dynamic model of enrollment, school choice, and achievement for Grades 4–9 in Mexico, finding that Prospera increases both enrollment and test scores, with telesecondary availability playing an important role. While their work emphasizes dynamic grade progression and longer-run program impacts, my analysis focuses on the single high-stakes transition into Grade 7 and explicitly models effort as an endogenous input, a key behavioral margin that their framework does not capture. This allows for a more detailed examination of how access constraints influence both enrollment and study behavior at this critical juncture. The opportunity cost channel is also relevant: Atkin (2016) and Shah and Steinberg (2017) find that higher outside wages reduce schooling and that work reduces learning. Consistent with these mechanisms, my estimates indicate that part-time work is strongly disfavored and substantially increases the marginal cost of effort.

A. Model Fit

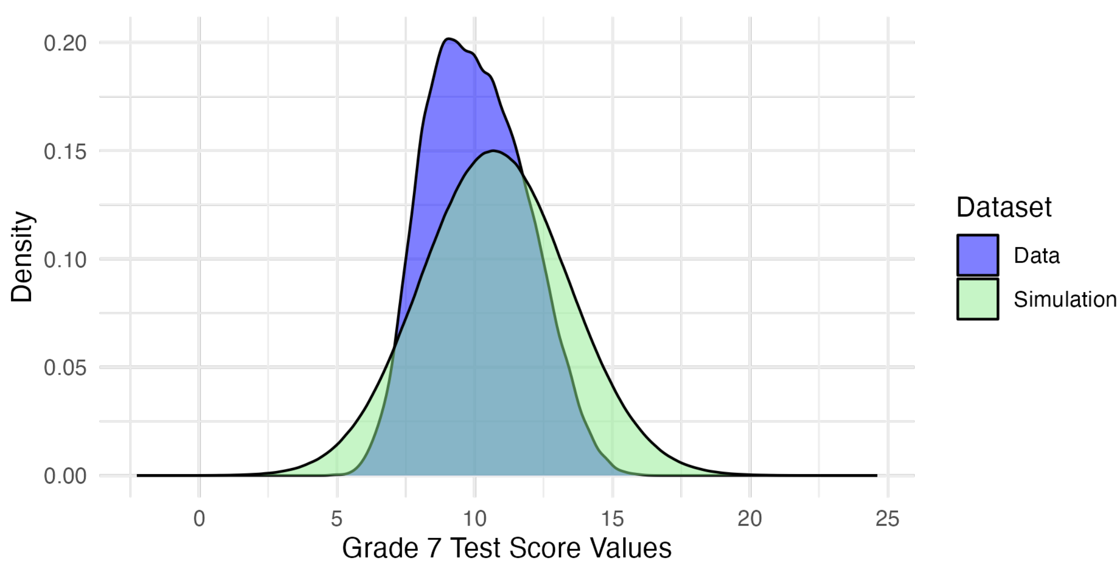
Before turning to the counterfactual simulations, it is important to assess how well the estimated model reproduces key patterns in the data. A close match between simulated and observed outcomes increases confidence that the model captures the main behavioral mechanisms and can generate credible predictions under alternative policy scenarios. This section compares outcomes simulated from the estimated model to those observed in the data. The estimation targets three main outcomes: test scores, effort, and school choice. Table 9 reports the means (and standard deviations, where applicable) for these outcomes. Overall, the fit is very good. The main exception is the test score distribution, which the model replicates less precisely (Figure 7). In contrast, the simulated effort distribution closely matches the data (Figure 8).

Table 9
Model Fit - 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		

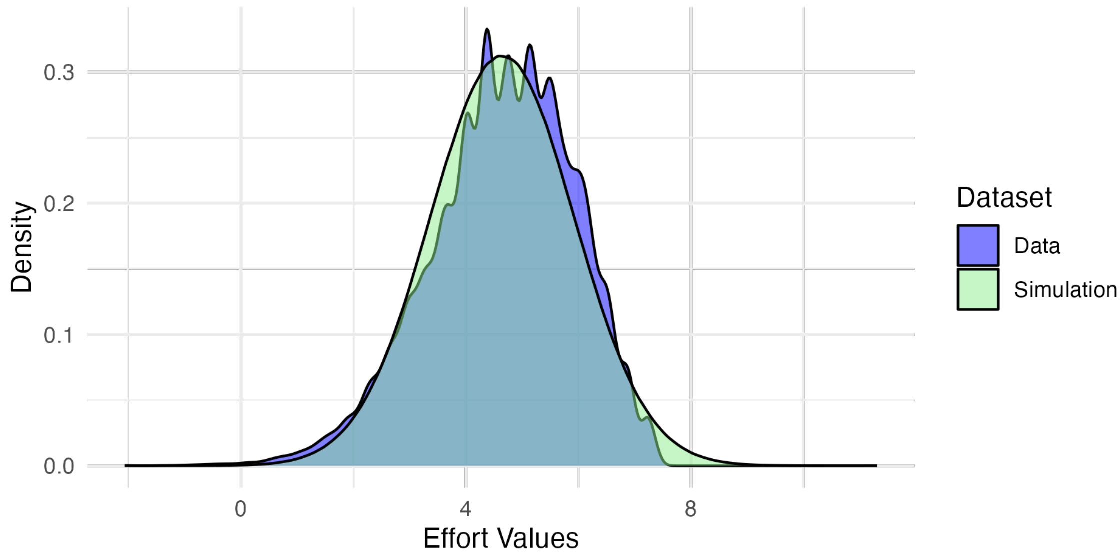
Note: The values in the "True Mean" and "True St.Dev." columns come directly from the data. The values in the "Simulated Mean" and "Simulated St.Dev" come from simulations using the parameter estimates.

Figure 7
Goodness of Fit: Test Score Distribution



Note: The blue histogram is the raw data and the green histogram is simulated data from the estimated model.

Figure 8
Goodness of Fit: Effort Distribution



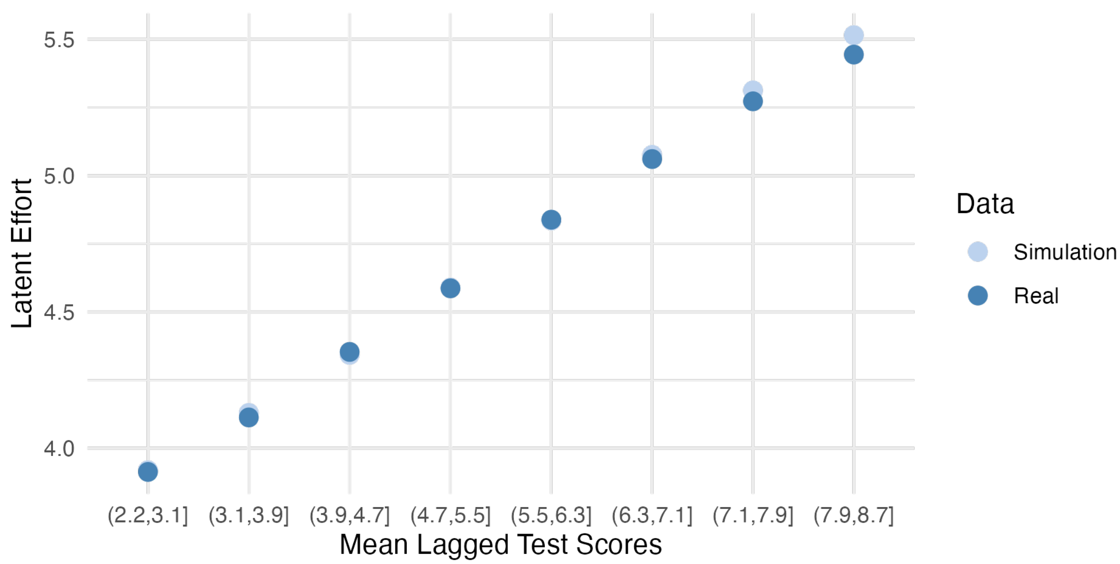
Note: The blue histogram is the raw data and the green histogram is simulated data from the estimated model.

Lagged test scores, treated as exogenous and serving as a proxy for ability, are strongly correlated in the data with both effort and dropout—two endogenous choices in the model. Figures 9 and 10 show that the model almost perfectly reproduces these relationships, including the non-linear link between lagged test scores and effort.

Finally, Figure 11 examines the relationship between school accessibility and dropout. Students are grouped into quintiles based on the distance to their nearest middle school, and mean dropout rates are calculated for each quintile. The model reproduces the observed U-shaped pattern but overestimates dropout among students whose nearest middle school is located in the same place as their primary school, as well as among those whose nearest middle school is very far away.

Taken together, these comparisons suggest that the model reproduces the key relationships among school access, labor participation, effort, and achievement with reasonable accuracy. While there are minor deviations in some subgroups, such as students with the closest or most distant schools, the overall fit supports using the estimated model to simulate how changes in policy or environment would affect enrollment, effort, and learning outcomes.

Figure 9
Goodness of Fit: Lagged Achievement and Effort



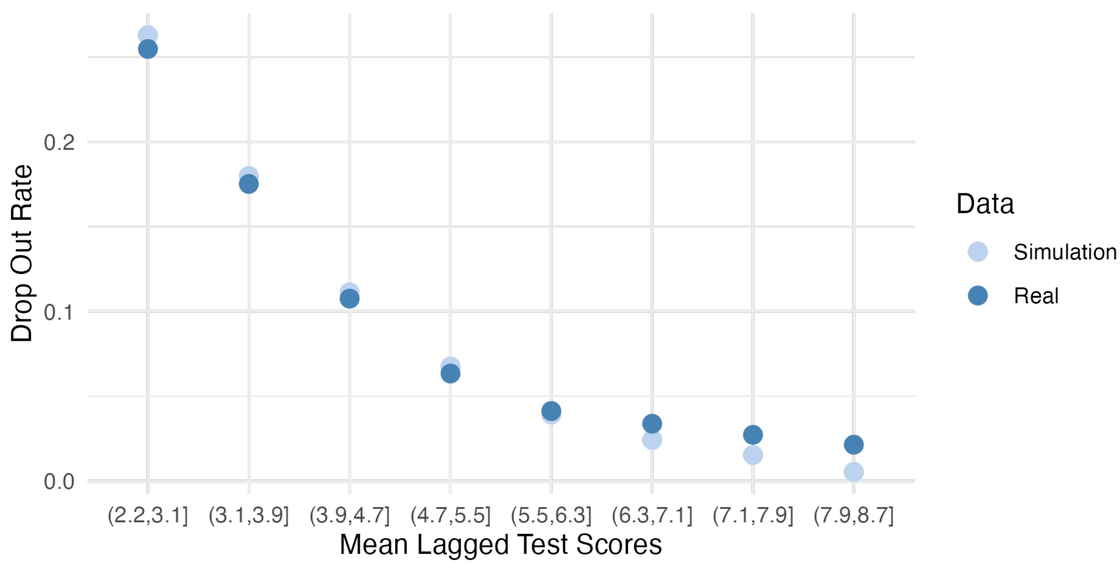
Note: The effort variable from the data and the effort variable generated by the model are plotted against average lagged test scores.

VI. Policy Simulations

With the estimated model, I evaluate a range of education policy scenarios that address different constraints students face when transitioning to middle school. These include enforcing existing child labor laws, expanding or re-targeting conditional cash transfers, improving school accessibility, and shifting perceptions of telesecundarias. The child labor enforcement scenario, which involves prohibiting all work for children younger than 14, serves as a benchmark for reducing dropout, but it would be costly and challenging to implement. I therefore consider alternative policies that could achieve similar reductions in dropout at potentially lower cost. Conditional cash transfers, which are already widely used in Mexico, can be effective, but require either higher benefit amounts or improved targeting of at-risk students. Improving physical access to schools by reducing travel costs, as well as changing perceptions of telesecundarias, also show substantial potential, with the latter suggesting a role for low-cost information campaigns.

To simulate these counterfactuals, I use the parameter estimates from the structural model to draw shocks and generate baseline choices for each student. For each policy

Figure 10
Goodness of Fit: Lagged Achievement and Drop Out



Note: The dropout decision from the data and the dropout decision generated by the model are plotted as a function of average lagged test scores.

scenario, I modify relevant variables (e.g., wages, transfer amounts, distance) or the set of feasible school and work options, then simulate choices again using the same shocks. Comparing baseline and policy simulations allows me to assess impacts on enrollment, achievement, work decisions, and the distribution of students across school types. I repeat each simulation 250 times and report average outcomes across runs, providing a consistent basis for comparing policy effects and, where possible, their costs.

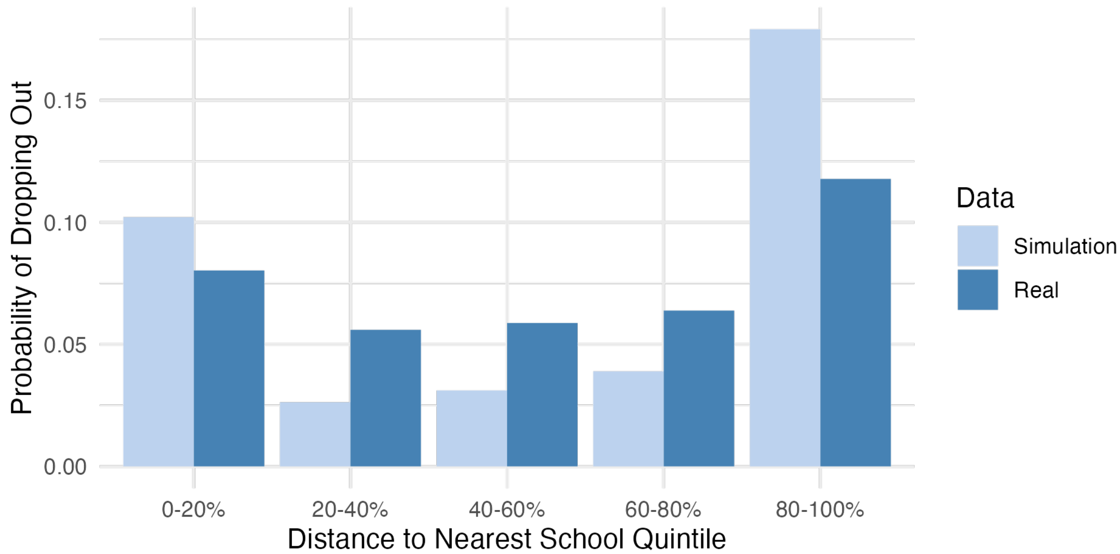
A. Ban Child Labor for Students Under 14

The first counterfactual considers the enforcement of existing child labor laws, prohibiting all work for children younger than 14. This policy removes both part-time and full-time labor options from students' choice sets. Students who had previously dropped out to work full time must now decide whether to enroll in school or remain out of school despite the absence of wage income. Students who had been working part time while enrolled are predicted to stay in school, as the value of their outside option declines more than the value of enrollment without work.

Table 10 shows that the dropout rate falls by 6.7 percentage points, though it remains

Figure 11

Goodness of Fit: Enrollment and Accessibility



Note: Each primary school is placed in a quintile based on the distance to the nearest middle school. The probability of dropping out is then calculated for each quintile, in the data and in the simulated data generated by the estimated model.

above 7%. Mean effort and test scores increase modestly overall. The effects are more pronounced for the 22% of students who had been working part time: as shown in Table 11, their effort rises by 5.9% of a standard deviation and their test scores by 4.4% of a standard deviation.

Table 10

Changes in Outcomes from Banning Child Labor

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

Note: The first counterfactual involves removing removing all labor options. Wages are set to zero, no part-time work is allowed, although students may still choose to drop out.

To understand why the reduction in dropout is limited, Table 12 compares students

Table 11
Banning Child Labor - Students Who Wanted To Work Part-Time

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

Note: The first counterfactual involves removing all labor options. Wages are set to zero, no part-time work is allowed, although students may still choose to drop out. The sample in this table includes all students who would like to work part-time in the baseline simulations, but stayed enrolled in school without working when the wages were set to zero.

who newly enroll under this policy to those who remain out of school. While the two groups differ somewhat in socioeconomic background, the largest difference is in school access: those who choose to enroll live, on average, nearly one kilometer closer to the nearest school. This underscores that prohibiting work does not eliminate the barriers posed by travel costs and access constraints.

Table 12
Banning Child Labor - Students Who Wanted To Work Full-Time

Background Variables	Enrolled	Stayed Out of School
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

Note: The first counterfactual involves removing removing all labor options. Wages are set to zero, no part-time work is allowed, although students may still choose to drop out. The sample in this table includes all students who would have liked to drop out and work full-time in the baseline simulations. In the counterfactual, some decided to enroll in school whereas others decided to stay out of school even without a wage.

Although banning all child labor may improve educational outcomes, such a policy would be costly and difficult to enforce in practice. This motivates the consideration of alternative policies that could improve enrollment in contexts where children continue to work. In Mexico, conditional cash transfers (CCTs) are a well-established instrument that can be adjusted in size or targeting to reach at-risk students. This approach is explored in the next counterfactual.

B. Increase Conditional Cash Transfer Benefits and Targeting

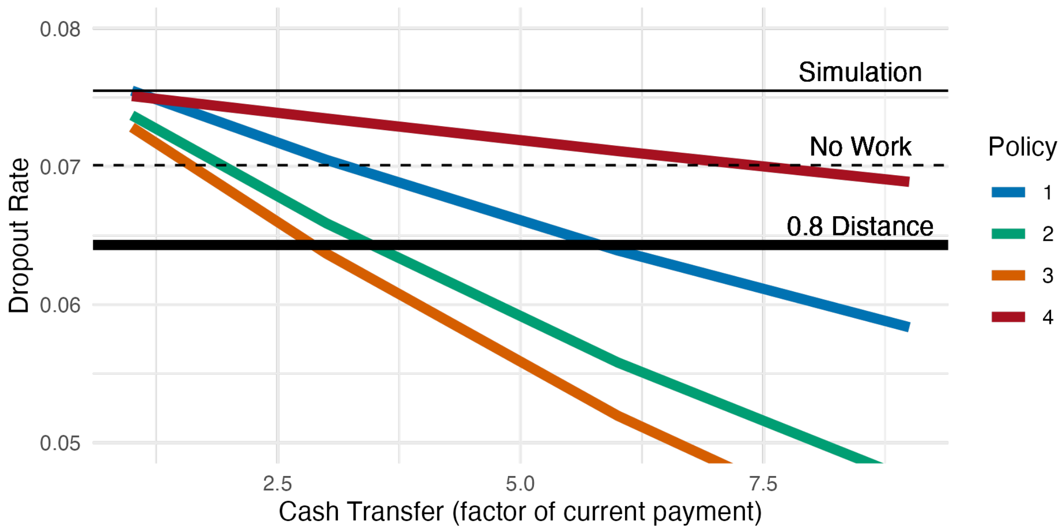
Figure 12 compares four alternative targeting strategies for conditional cash transfers (CCTs), with transfer amounts ranging from the 2010 Prospera level up to nine times that value. The x-axis shows the transfer size and the y-axis shows the resulting dropout rate. Three horizontal reference lines are included: the "Simulation" line shows the baseline dropout rate, the "No Work" line shows the dropout rate under a full child labor ban, and the "0.8 Distance" line shows the dropout rate when all travel distances are reduced by 20%, highlighting the strong influence of school access.

The four policies differ in who is eligible for the transfer. Policy 1 increases payments to the current Prospera beneficiaries, making it the simplest to implement. Policy 2 extends eligibility to these same beneficiaries plus students from families below the median income. Policy 3, a hypothetical upper-bound scenario, offers the transfer to all students who would drop out in the baseline simulation. While operationally infeasible, this scenario shows the maximum potential impact of targeting the transfer perfectly. Finally, Policy 4 offers the transfer to all students whose nearest middle school is more than two kilometers from their primary school, focusing on poor school access as a key barrier to enrollment.

The results show that increasing the transfer amount can be a highly effective way to reduce dropout rates, but the cost can be substantial. At transfer levels close to the current value, expanding eligibility to additional low-income families (Policy 2) has little effect on dropout. At higher transfer amounts, however, broader targeting produces much larger reductions. Cost-effectiveness patterns, reported in Table 13, paint a different picture. While Policy 3 achieves the largest reduction at the lowest cost per dropout averted, it is not operationally realistic. Among feasible options, Policy 4 is most cost-effective, suggesting that targeting CCTs toward students with poor school access could be a promising approach. This finding also points to the potential benefits of complementary policies such as subsidized transportation in remote areas.

Figure 12

Counterfactual Policy: Dropout Rate and Conditional Cash Transfers



Note: 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 dashed horizontal "No Work" line represents the dropout rate from banning child labor. The solid "0.8 Distance" line represents the dropout rate from decreasing all travel distances by 20%.

Table 13

Counterfactual CCT Policies - Comparing Costs

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

Note: Comparing costs for different CCT policies to reduce the drop out rate to the "No Work" level. The transfer amounts are in factors of the 2010 transfers. The total costs are also in factors of the actual 2010 total cost.

C. Improve Perceptions of Telesecundarias

Students most at risk of dropping out are typically in rural areas, where their nearest middle school is often a telesecundaria. The parameter estimates indicate that these schools are perceived less favorably than general schools, despite recent evidence that their educational quality is comparable to local alternatives (Borghesan and Vasey 2024). This counterfactual examines the effect of eliminating that perception gap by valuing telesecundarias the same as general schools. As shown in Table 14, the results are striking: the dropout rate would fall to 5%, and the share of students enrolled in telesecundarias would rise from 19% to 26%. Test scores, effort, and the share of students working while enrolled remain essentially unchanged. Given the growing evidence on the effectiveness of telesecundarias, an information campaign aimed at improving perceptions could be a relatively low-cost intervention with potentially large benefits.

Table 14
Changing Perceptions Around Telesecundarias

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

Note: Changes in outcomes if Telesecundarias were valued the same as General schools.

VII. Conclusion

Increasing human capital is widely regarded as one of the most effective ways for developing countries to promote economic growth and reduce inequality. Ensuring that all children attend school to at least a minimum age and receive a high-quality education is therefore a central policy priority. In many low- and middle-income countries, however, the persistence of child labor complicates these goals, both by drawing children out of school and by reducing the time and energy they can devote to learning. While the existing literature on school choice is extensive, few frameworks explicitly integrate child labor and its interaction with educational decisions. This paper addresses that gap by jointly modeling schooling and labor choices and by incorporating a mechanism through which work affects achievement: the study effort that children allocate to their education.

I develop and estimate a random utility model over discrete school–work alternatives in which study effort is determined as the solution to an optimization problem for each option. Students who do not enroll in school are assumed to work full-time and receive the associated wage. Students who enroll may also choose part-time work, gaining additional income but facing a higher marginal cost of effort. Effort, in turn, enters a value-added achievement equation, so that students who work and therefore exert less effort achieve lower test scores than they would otherwise.

The estimation combines several rich data sources: administrative records on nationwide standardized tests in math and Spanish, survey data from students, parents, and principals, geocoded school locations, and Mexican census data on local labor market wages and hours worked. Most model parameters are precisely estimated, allowing for credible counterfactual simulations.

The policy analysis shows that banning all child labor increases effort and achievement, with part-time workers benefiting the most, and reduces dropout by 6.6%. However, the dropout rate remains above 7%, and the policy would be costly and difficult to enforce. Increasing the magnitude and improving the targeting of conditional cash transfers can achieve similar reductions in dropout by encouraging enrollment rather than prohibiting work, though higher transfers substantially raise program costs. Finally, policies that improve school access or shift perceptions of telesecundarias, particularly in rural areas, emerge as especially promising, though they are more speculative and require further evidence on feasibility and cost-effectiveness.

By integrating school choice, labor supply, and effort into a unified framework, this paper offers a richer understanding of the constraints shaping educational investment in settings where child labor is prevalent. The results highlight that effective policy design must address both the opportunity costs of schooling and the perceived quality of available options.

References

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

Alcaraz, Carlo, Daniel Chiquiar, and Alejandrina Salcedo. 2012. "Remittances, schooling, and child labor in Mexico." *Journal of Development Economics* 97 (1): 156–165.

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, and Tristan Zajonc. 2011. "Do value-added estimates add value? Accounting for learning dynamics." *American Economic Journal: Applied Economics* 3 (3): 29–54.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. 2008. "A dime a day: The possibilities and limits of private schooling in Pakistan." *Comparative Education Review* 52 (3): 329–355.

Atkin, David. 2016. "Endogenous skill acquisition and export manufacturing in Mexico." *American Economic Review* 106 (8): 2046–2085.

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, Susan W Parker, Petra Todd, and Weilong Zhang. 2025. "Prospering through Prospera: A dynamic model of CCT impacts on educational attainment and achievement in Mexico." *Quantitative Economics* 16 (1): 133–183.

Boardman, Anthony E, and Richard J Murnane. 1979. "Using panel data to improve estimates of the determinants of educational achievement." *Sociology of Education*, pp. 113–121.

Borghesan, Emilio, and Gabrielle Vasey. 2024. "The marginal returns to distance education: evidence from Mexico's telesecundarias." *American Economic Journal: Applied Economics* 16 (1): 253–285.

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.

Cummings, Patricia Murrieta. 2016. "Child labor and household composition: Determinants of child labor in Mexico." *Asian Journal of Latin American Studies* 29 (3): 29–54.

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.

Duflo, Esther. 2001. "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment." *American economic review* 91 (4): 795–813.

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.

Hanushek, Eric A. 1979. "Conceptual and empirical issues in the estimation of educational production functions." *Journal of human Resources*, pp. 351–388.

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*.

Kozhaya, Mireille, and Fernanda Martinez Flores. 2022. "School attendance and child labor: Evidence from Mexico's Full-Time School program." *Economics of Education Review* 90:102294.

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.

Muralidharan, Karthik, and Nishith Prakash. 2017. "Cycling to school: Increasing secondary school enrollment for girls in India." *American Economic Journal: Applied Economics* 9 (3): 321–350.

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.

Shah, Manisha, and Bryce Millett Steinberg. 2017. "Drought of opportunities: Contemporaneous and long-term impacts of rainfall shocks on human capital." *Journal of Political Economy* 125 (2): 527–561.

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.

Tagliati, Federico. 2019. "Child labor under cash and in-kind transfers: evidence from rural Mexico."

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.

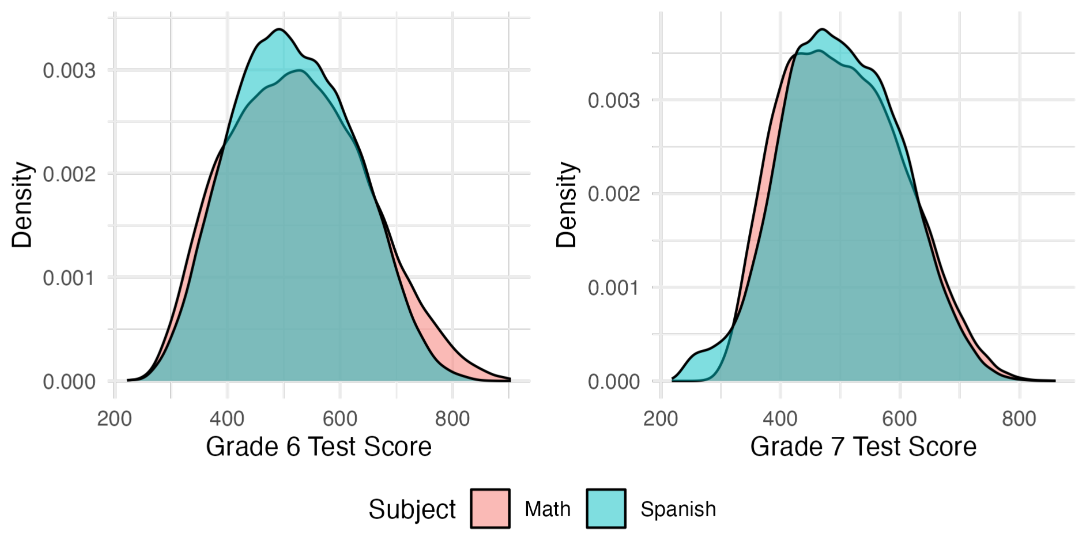
Todd, Petra E, and Kenneth I Wolpin. 2003. "On the specification and estimation of the production function for cognitive achievement." *The Economic Journal* 113 (485): F3–F33.

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
Raw Test Score Distributions



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 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 shown in Table B-2.

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 1 shows a histogram of the final effort measures.

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

C. Estimation Strategy Details

1. Guess parameters. There are 30 parameters in this version of the model:
   - 11 coefficients for each of the achievement value added equation
   - 1 parameter for the variance of the error term for the value added equation
   - 16 coefficients in the utility equation
   - 1 parameter for the variance of the effort distribution
   - 1 parameter for the scale parameter in the multinomial logit
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 test score 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 **??**.
   - 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 the 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, and their responses to the effort questions.

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. 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.

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%.

- 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*