The Marginal Returns to Distance Education: Evidence from Mexico's Telesecundarias

By EMILIO BORGHESAN AND GABRIELLE VASEY*

This paper analyzes a large-scale and long-running distance education program in Mexico. We estimate marginal treatment effects (MTEs) for learning in math and Spanish in telesecundarias relative to traditional Mexican secondary schools using an empirical framework that allows for unobserved sorting on gains. The estimated MTEs reveal that school choice is not random and that the average student experiences significant improvements in both math and Spanish after just one year of attendance in telesecundarias. We find that the existing policy reduces educational inequality, and our policy-relevant treatment effects show that expanding telesecundarias would yield significant improvements in academic performance.

JEL Codes: I21, I24, O15.
Keywords: Education Policy, Achievement Gap, Education and Inequality

Distance education elicits strong reactions from parents, policymakers, and teachers alike (Almagor, 2021). Recent studies of the effects of pandemic-related school closures on academic achievement in developed countries show that more time away from in-person school is associated with less learning (Engzell, Frey and Verhagen, 2021; Maldonado and De Witte, 2021). The effectiveness of distance education depends on two factors: how it is implemented and the quality of the counterfactual alternative. Distance education can be implemented as an online version of in-person schooling or it can be scaled, with a small number of highly effective communicators delivering lectures to multiple school districts. In addition, remote learning is more likely to generate learning gains where the quality of in-person alternatives is low. These two factors mean that distance education may be of particular value in developing countries, where rural-urban divides and high rates of teacher absenteeism deprive many students of high-quality instruction (Duflo, Hanna and Ryan, 2012). Yet, large-scale implementation of a

* Borghesan: University of Pennsylvania; email: borghesa@sas.upenn.edu. Vasey: Concordia University; email: gabrielle.vasey@concordia.ca. We would like to thank the editor and three anonymous referees. We also thank Jere Behrman, Susan Parker, Uta Schönberg, and Petra Todd for their help and guidance, as well as seminar participants at the University of Pennsylvania, LACEA/LAMES, the Young Economists Symposium, Society of Labor Economics, Canadian Economic Association, Université du Québec à Montréal, and Université de Montréal for helpful comments. We are also deeply thankful to Mira Potter-Schwartz, Elizaveta Brover, and Erika Trevino Alvarado for providing excellent research assistance. The Mexican Secretariat of Public Education (SEP) has generously provided data for this analysis. We thank Hector Robles Vaquez for preparing databases and for assistance working with these data. We are grateful for financial support from NSF award #1948943 and from the University of Pennsylvania School of Arts and Sciences internal grants "Making a Difference in Diverse Communities" and the "Dean's Global Inquiries Fund".

single remote curriculum is rare and there has been little research in economics evaluating the effectiveness of such programs.

In this paper, we study the effects of a long-running distance education program on student learning in Mexico. Since 1968, Mexico has undertaken an ambitious effort to provide a distance education option to secondary school students.[1] These students decide between traditional schools, with subject-specific teachers delivering their own lectures in person, and telesecundarias, brick and mortar establishments where students watch televised lectures and work on standardized assignments under the supervision of a single adult monitor. If shown to be effective, telesecundarias could provide a model for operating distance education at scale that differs considerably from the pandemic approach to schooling.

We analyze the effectiveness of telesecundarias relative to traditional public schools in Mexico using data on low-stakes end-of-year standardized tests in math and Spanish—the Evaluación National de Logro Académico en Centros Escolares, or ENLACE.[2] We examine students who were in the sixth grade in 2007/08 and who advanced to secondary school the following year. Figure 1 presents differences in mean scores on the ENLACE exams between telesecundaria and traditional secondary students for this sample of public school students for grades six through nine. The figure shows that in grade six, the year prior to enrollment in secondary school, students who will attend telesecundarias the following year score 46.7 points lower in math and 50.5 points lower in Spanish, on average, than their peers who will eventually attend traditional secondary schools. These gaps in the sixth grade, of 0.387 and 0.471 standard deviations, respectively, diminish considerably after students enroll in secondary school. For each year of enrollment, the performance gap between students at traditional schools and telesecundarias narrows in Spanish, and in math students at telesecundarias overtake their peers at traditional schools by grade eight and are performing 28.2 points better by grade nine.

The marked improvement of telesecundaria students relative to their traditional school peers is neither caused by differential rates of dropout, as we restrict our sample to students who advance to secondary school, nor is it likely to be an artifact caused by differential rates of cheating or exam-sitting (see Appendix A). This improvement is also robust to the inclusion of a large number of controls for family background and geographic location. Nevertheless, since secondary school type is not randomly assigned, the extent to which this improvement represents a causal effect of telesecundarias or whether it stems from selection

[1]Secondary schools in Mexico enroll students in grades seven through nine and are akin to middle schools in the United States, while upper secondary schools enroll students in grades ten through twelve. Throughout the paper, we use the terms "secondary" and "upper secondary" to be consistent with the Mexican educational system.

[2]Numerous studies have used the ENLACE exams as a means of evaluating educational interventions (De Hoyos, Garcia-Moreno and Patrinos, 2017; Dustan, De Janvry and Sadoulet, 2017; Estrada and Gignoux, 2017; Avitabile and De Hoyos, 2018; Estrada, 2019; Michaelsen and Salardi, 2020). Scores on these exams have been shown to predict important life outcomes including university enrollment and wages (de Hoyos, Estrada and Vargas, 2021).

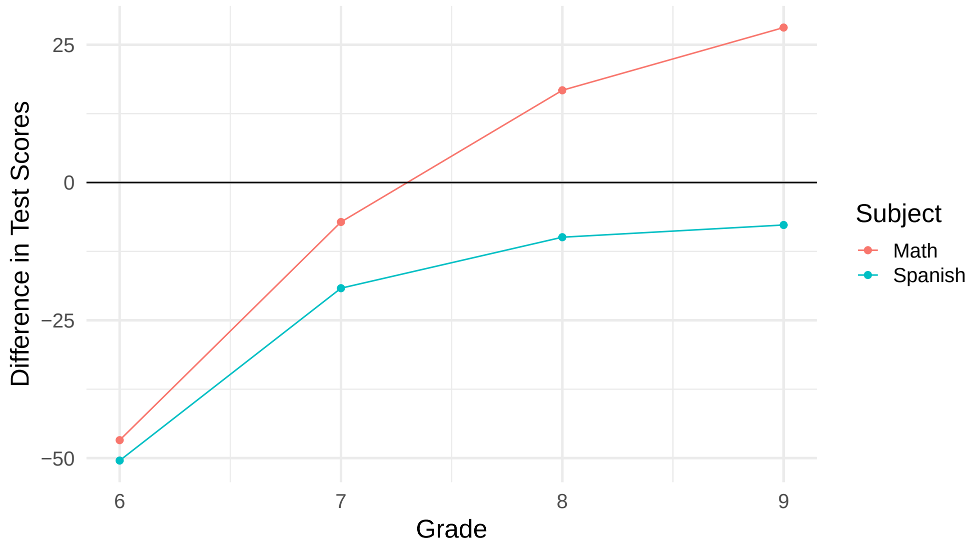


FIGURE 1. DIFFERENCES IN MEAN TEST SCORES BETWEEN TELESECUNDARIAS AND TRADITIONAL SCHOOLS

Note: The figure shows the difference in mean scores on the math and Spanish ENLACE exams between telesecundaria and traditional public school students in Mexico. Students in grade 6 attend primary school. Starting in grade 7, public school students attend either traditional secondary schools or telesecundarias. The standard deviations on the math ENLACE exams are 120.7, 100.7, 104.0, and 119.3 in grades 6, 7, 8, and 9, respectively. The standard deviations on the Spanish ENLACE exams are 107.3, 100.0, 109.2, and 100.8 in grades 6, 7, 8, and 9, respectively.

based on unobservable characteristics is unclear.

To assess the causal effect of telesecundarias on student learning, this paper combines a value-added model with the marginal treatment effects (MTE) framework, introduced by Björklund and Moffitt (1987) and generalized in Heckman and Vytlacil (1999; 2001*b*; 2005; 2007). We measure learning in two subjects, math and Spanish, by value added, the intertemporal difference in scores on the ENLACE exams administered the year before and one year after the start of secondary school. We focus on the transition from primary school to secondary school (grades six to seven), as this is when students face a choice regarding the type of school to attend. We define the relative effectiveness of telesecundarias as the difference in value added between the two school types between grades six and seven. Focusing on value added allows us to isolate how much student knowledge in seventh grade is directly attributable to the type of secondary school attended.

We use the MTE framework to jointly model students' choice of school and their academic outcomes at each school. This framework, which allows for the possibility that the choice of secondary school may be correlated with unobserved factors affecting performance on the exams, is a flexible semiparametric selection model that can estimate treatment effects for policies as they are currently implemented as well as evaluate a wide range of counterfactual policies. It is important to conduct the MTE analysis that we do in this paper instead of using traditional instrumental variables (IV) methods for several reasons. The estimates that are obtained by IV with a continuous instrument can be represented as a particular weighted average of the MTE and correspond to some treatment effect, but they do not usually identify typical treatment parameters of interest in the presence of sorting on gains. We use our estimated MTEs to calculate these parameters – the average effect of treatment (ATE), the average effect of treatment on the treated (TT), and the average effect of treatment on the untreated (TUT) – as well as analyze additional counterfactual policies that expand access to telesecundarias. Finally, we use the MTE to analyze how telesecundarias influence the rural-urban test score gap and the gap in test scores between students who speak an indigenous language at home and those who speak Spanish. The results in this paper highlight the flexibility and usefulness of the MTE framework as a tool for policy evaluation.

Identification of the MTE requires an instrumental variable that affects the decision of which school to attend but does not directly affect outcomes at each school. The instrument we use in this paper is a measure of relative distance, which we compute as the difference between two distance measures. The first measure is the distance between each student's primary school and the nearest telesecundaria, while the second measure is the distance between the same primary school and the nearest traditional secondary school. The resulting instrument measures how many kilometers farther the telescundaria is compared to the traditional school.[3,4]

This instrumental variable, relative distance, is highly predictive of attendance in telesecundarias: A one kilometer reduction in this measure, meaning that the telesecundaria becomes relatively closer, causes a student's probability of attending a telesecundaria to increase by 2.93 percentage points on average. Cameron and Taber (2004) and Carneiro and Heckman (2002) have raised concerns that distance to secondary school is correlated with academic ability in the United States. We discuss why endogeneity of this sort is less likely in Mexico than in the US, and we conduct balancing tests to show that, after including a range of controls, the instrument is uncorrelated with measures of cognitive skills that we observe in the data.

We find that telesecundarias are highly beneficial: The average treatment effect (ATE) of telesecundaria attendance relative to attendance in traditional schools is a 35.0 point increase in math scores (equivalent to 0.348 standard deviations on the exam) and a 22.5 point increase in Spanish scores (equivalent to 0.225 standard

[3] A large body of empirical work uses a measure of distance to school as an instrument. See Card (1995), Kane and Rouse (1995), Kling (2001), Currie and Moretti (2003), Cameron and Taber (2004), Carneiro, Heckman and Vytlacil (2011), and Carneiro, Lokshin and Umapathi (2017).

[4] We eliminate students from the sample who lack both schooling options within a 20 km radius.

deviations) after just one year of attendance.[5] These ATEs conceal considerable heterogeneity in who benefits from telesecundarias. Some students see gains of close to 50 points, while others experience no benefit. Our findings suggest that much of the improvement of telesecundaria students documented in Figure 1 represents causal effects of telesecundarias.

Our analysis uncovers a pattern of negative sorting on gains. The students who would benefit the most from attending telesecundarias are those who are least likely to attend. We find that sorting is primarily driven by variability in outcomes at traditional schools, while outcomes at telesecundarias are more homogeneous, consistent with these schools offering the the same (televised) lectures across schools.

We use our estimated MTEs to evaluate counterfactual policies that expand access to telesecundarias. The first policy we consider is a dramatic increase in telesecundaria availability that reduces the instrument, relative distance, by five kilometers (km) for everyone in the sample. The second policy is a school building program that constructs a telesecundaria adjacent to the twenty-two percent of Mexican primary schools without one within five km. We find that the first policy raises math (Spanish) scores by 37.3 (25.0) points, while the second raises scores by 23.6 (17.7) points. The effects of the two policies differ and neither corresponds to the estimates obtained by a Two-Stage Least Squares regression that uses distance as an instrument (30.4 and 18.7 points for math and Spanish, respectively), highlighting the importance of adopting a framework allowing for heterogeneous treatment effects and self-selection as we do in this paper. These policy-relevant treatment effects show that further expansions of this already widespread program would cause gains in test scores.

We additionally use the MTE framework to estimate the causal effects of telesecundarias on two prominent divisions in Mexico: the gap in test scores between rural and urban students and between students who speak Spanish at home and those who speak an indigenous language. Our findings indicate that the rural-urban test score gap would be 128% larger in math and 43% larger in Spanish if all students were to attend traditional schools. Similarly, the Spanish-Indigenous language test score gaps would be 20% larger in math. Our findings suggest that, not only do telesecundarias improve educational outcomes in Mexico, they do so in a way that reduces persistent educational inequities across different subgroups of the population.

Our paper contributes to the growing literature on how education technology can be used in low- and middle-income settings.[6] Beg et al. (2022) and Johnston and Ksoll (2017) study the effects of video lectures and remote instruction rela-

[5]As both the mean and standard deviation of test scores in the seventh grade may be influenced by the treatment effect, we conduct our analysis using raw scores on the exams. The standard deviations on the seventh grade ENLACE exams (100.7 for math and 100.0 for Spanish) are close enough to 100 to allow for an easy translation between points and standard deviations.

[6]Earlier work on the interaction of technology and education is surveyed in three excellent reviews by Bulman and Fairlie (2016), Escueta et al. (2017), and Rodriguez-Segura (2020).

tive to traditional in-person schooling in Pakistan and Ghana using randomized controlled trials (RCTs). Bianchi, Lu and Song (2022) study a large scale distance-learning intervention in China similar to the one in this paper. They find average improvements in math and Chinese scores of 0.18 and 0.23 standard deviations, which are of a similar magnitude to our findings. Fabregas (2020) and Navarro-Sola (2021) study the impacts of telesecundarias on labor market outcomes and find that they raise educational attainment and future income. Behrman, Parker and Todd (2020) and Behrman et al. (2021) instead study the interaction between telesecundarias and conditional cash transfers. Our contribution to the literature on telesecundarias is through demonstrating their significant causal effects on student learning. Our findings suggest that the effects of telesecundarias on boosting earnings and participation in the formal sector documented in prior work result not only from educational attainment but also from improved academic performance conditional on staying in school.

A second related strand of literature evaluates the effects of schooling using the MTE framework. Carneiro, Heckman and Vytlacil (2011) estimate positive treatment effects of college attendance on income and uncover a pattern of positive sorting on gains. Carneiro, Lokshin and Umapathi (2017) find similar results for the returns to attending secondary school in Indonesia using distance to the nearest secondary school as an instrument. Cornelissen et al. (2018) analyze the decision of parents to enroll their children in daycare in Germany, and, unlike the previous two papers, find a pattern of negative sorting on gains. Students who are not enrolled would experience increases in their readiness for primary school had they attended child care, while those currently attending experience little benefit. Walters (2018) finds reverse sorting on gains into charter schools in Boston, as do De Groote and Declercq (2021) when evaluating elite high schools in Flanders (Belgium). An advantage of our paper relative to this strand of literature is that we are able to nonparametrically identify the MTE over its full support, and we estimate all our policy effects without strong assumptions on the joint distribution of unobservables or on the shape of the MTE.[7] The approach provides assurance that our estimated treatment parameters are not biased by a subpopulation with markedly different treatment effects who are especially likely (or unlikely) to attend telesecundarias.

This paper proceeds as follows. Section I provides information on secondary schooling in Mexico, while section II describes the data. Section III describes the model and empirical strategy, and section IV presents our main empirical results. Section V presents estimates of counterfactual treatment effects and explores telesecundarias' effects on educational inequality. Section VI concludes.

[7]Sapelli and Vial (2002) conduct a similar analysis of the effects of private schools on test scores in Chile. They estimate treatment effects using a generalized Roy Model, which assumes that unobservables have a multivariate normal distribution, without estimating the MTE curve directly.

I. Secondary Schooling in Mexico

Throughout the twentieth century, Mexico struggled to attract qualified teachers to rural areas to instruct the millions of school-age children living there. Telesecundarias were introduced in 1968 as a solution to this problem. Telesecundarias are physical structures with four key scholastic components. The first is the television program. Every subject begins with students watching fifteen minutes of a pre-recorded televised lecture. Lectures for each subject are recorded in Mexico City by high-quality instructors, so-called *telemaestros*, who are selected for their communication skills and undergo extensive training. Considerable effort is made to ensure that the videos are of high quality: The typical fifteen-minute televised lecture undergoes twenty days of production time and costs $30,000-$50,000 (Calderoni, 1998).

Following the video lectures, a single teacher leads students in a 35-minute lesson on the same subject. The teacher does not specialize in a particular subject: Students have the same teacher for all courses. This teacher follows a guide designed for telesecundaria teachers that is filled with suggestions for each subject. The 35 minutes are spent in myriad ways, with the teacher leading question and answer sessions, engaging students in group activities, and giving assignments for students to do on their own.

The third educational resource is an encyclopedia-like book that contains the essential information taught in each subject. These books are similar to textbooks in traditional secondary schools and are used by students as references while doing their assignments. The final component of telesecundaria education is the learning guide. Like a workbook, learning guides are filled with questions that students can answer individually as well as suggestions for group activities that reinforce learning. Class time is frequently devoted to doing assignments in the learning guides.

The four main educational components – televised instruction, in-person teaching, reference texts, and learning guides – are designed to be complementary. Ethnographic research indicates that students see each component as reinforcing the knowledge acquired through the televised lectures (Estrada, 2003).

Telesecundarias were first introduced in rural areas, predominately in Mexico's poorer South. While they have expanded into suburban and urban areas, students from the South and from rural areas are still over-represented (see Table 1). The typical telesecundaria is a purpose-built structure with between three and nine classrooms, a library, restrooms, a science lab, and a playground. Students in telesecundarias and traditional schools attend school for the same number of hours per week (30) and days per year (200). A reform in 1993 mandated schooling through grade nine and resulted in increases in both the construction of new telesecundarias and telesecundaria enrollment. We study the cohort of students who were in grade six in 2007/08, after secondary schooling became compulsory.

II. Data

We combine data on students from several sources. The first is an administrative data set comprising the universe of student scores on the ENLACE exams in math and Spanish (SEP, 2018*b*). The ENLACE exams, administered to Mexican schoolchildren in grades three through nine between 2006 and 2013, test end-of-year academic knowledge in math, Spanish, and a rotating third subject. They are low-stakes exams, with no bearing on a student's GPA, graduation, or admission to higher education. The exams were designed to have a mean of 500 and a standard deviation of 100 in their first year of implementation, but exam scores were not re-standardized in subsequent years. They are scored using item response theory, which allows improvements in test scores to be interpreted as learning gains. In addition to test scores, the ENLACE data contains information on the age, gender, conditional cash transfer status, school attendance, school ID, and school type for each student.

We link the ENLACE data with information on student, parent, and school characteristics from a random sample of schools (SEP, 2018*a*). These surveys provide detailed information on parental education, monthly family income, home infrastructure, number of siblings, and other household characteristics. The combination of the ENLACE data with the secondary school surveys produces an exceptionally rich data set that is representative of the population of Mexican schoolchildren.[8]

We combine these two data sources with information on the latitude and longitude of each primary and secondary school and calculate the distance in kilometers between the primary school each student attends and the nearest secondary school of each type (Escuelasmex, 2018). We subtract the distance to the nearest traditional school from the distance to the nearest telesecundaria to obtain a measure of relative distance. We use the relative distance measure as an exclusion restriction that affects the choice of school but does not affect test scores directly. While it might be preferable to measure distance from the student's actual home (rather than primary school) to each secondary school, this would rely on information that is not present in any of our sources. In the next section, we show that the instrument we construct is highly predictive of attendance in telesecundarias.

Figure 2 presents the distribution of relative distance by school attendance. A negative value on the x-axis indicates that a telesecundaria is closer, while a positive value indicates that a traditional school is closer. The figure reveals that students mostly attend the school that is closer, but when the two schools are equally close (relative distance $= 0$), many more students attend traditional schools.

We impose minimal sample selection criteria. First, we eliminate the roughly

[8]Both the test score data and the survey data were administered by the Mexican Secretariat of Public Education (SEP), who provided our research team with access to the data.

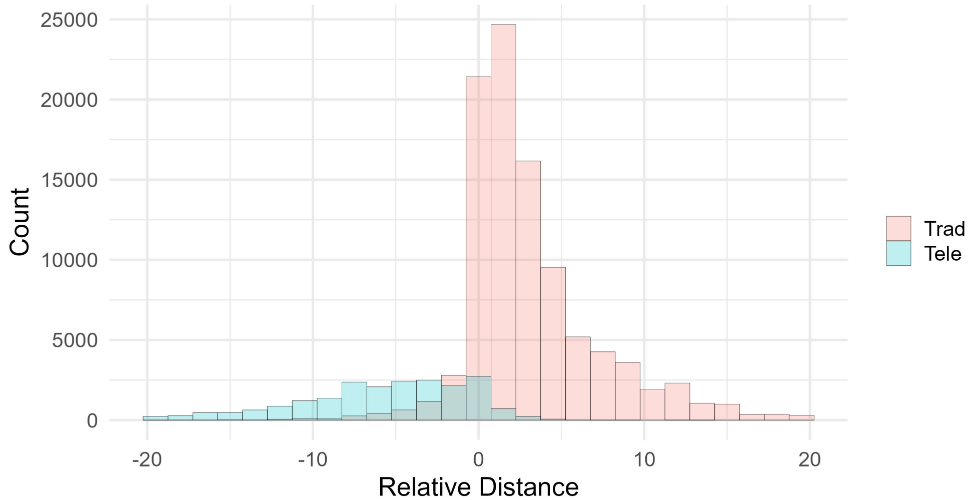


FIGURE 2. RELATIVE DISTANCE TO TELESECUNDARIA

Note: The figure plots a histogram of the instrumental variable by treatment status. Trad refers to students attending traditional schools, while Tele refer to students in telesecundarias. The instrument is the difference between two measures of distance. The first is the distance from the student's primary school to the nearest telesecundaria, while the second is the distance from the student's primary school to the nearest traditional school. Relative distance is negative whenever telesecundarias are closer.

seven percent of sixth grade students who are unobserved in the seventh grade.[9] We also eliminate the nine percent of students who attend private schools. There is little overlap in the distributions of family income for students who attend private schools versus telesecundarias, so we do not consider private school to be a feasible alternative for the vast majority of Mexican schoolchildren who are considering attending a telesecundaria. We remove students from three states – Guerrero, Michoacan, and Oaxaca – that have limited data on ENLACE exam scores. Administration of the ENLACE exams was opposed by the teachers unions in these states.

Finally, we omit from our analysis students whose relative distance measure lies outside the middle 99% of the distribution and students who attend a secondary school more than twenty km from their primary school. We want to consider students who have a choice set consisting of two feasible alternatives, and so we drop students with only one nearby school. We end up with a sample of $118,526$ students.[10]

[9]Most of these students drop out between the sixth and seventh grade, but a small number are unobserved in grade 7 and then observed in later grades. Section V discusses how omitting dropouts influences the interpretation of treatment effects corresponding to counterfactual policies that increase access to telesecundarias.

[10]See Appendix F for complete details on sample selection.

TABLE 1—SUMMARY STATISTICS

	Telesecundaria		Traditional	
	Mean	S.D.	Mean	S.D.
Math Score (7th grade)	499	(98.1)	504	(98.1)
Spanish Score (7th grade)	491	(96.1)	506	(96.7)
Math Score (6th grade)	491	(111)	536	(118)
Spanish Score (6th grade)	485	(95.1)	532	(102)
Relative Distance	-5.65	(4.97)	3.09	(3.97)
Age	12.2	(0.800)	11.9	(0.599)
Siblings	3.61	(2.38)	2.38	(1.65)
Prospera	0.621	(0.485)	0.153	(0.360)
Female	0.505	(0.500)	0.508	(0.500)
Computer at Home	0.131	(0.338)	0.422	(0.494)
Rural Residence	0.675	(0.469)	0.086	(0.280)
Northern State	0.061	(0.239)	0.294	(0.455)
Books in the Home : ≤ 10	0.713	(0.452)	0.468	(0.499)
Books in the Home : 20	0.179	(0.383)	0.260	(0.439)
Books in the Home : 50	0.064	(0.245)	0.157	(0.364)
Books in the Home : ≥ 100	0.043	(0.203)	0.115	(0.319)
Mother's Education : Primary	0.741	(0.438)	0.408	(0.491)
Mother's Education : Middle	0.204	(0.403)	0.284	(0.451)
Mother's Education : Secondary	0.042	(0.201)	0.237	(0.425)
Mother's Education : Postsecondary	0.012	(0.109)	0.071	(0.257)
Income (Pesos/mo) : ≤ 2500	0.559	(0.497)	0.224	(0.417)
Income (Pesos/mo) : 2500-2999	0.276	(0.447)	0.302	(0.459)
Income (Pesos/mo) : 3000-7499	0.126	(0.332)	0.320	(0.466)
Income (Pesos/mo) : ≥ 7500	0.039	(0.194)	0.155	(0.362)
Observations	20809		97717	

Note: The table displays summary statistics for outcome variables, covariates, and the instrument – relative distance – for students in our estimation sample. The statistics are broken down by the type of secondary school attended.

Table 1 presents summary statistics for the outcome variables, covariates, and the instrumental variable (relative distance) by school type. The table reveals that students who attend telesecundarias are disadvantaged according to a wide range of measures relative to students who attend traditional schools. They are disproportionately beneficiaries of the conditional cash transfer program Prospera, they come from poorer household with less educated mothers, and they fare worse academically in the year prior to secondary school.[11] Nevertheless, they make up over half of the gap in Spanish and nearly the entire gap in math relative to their peers at traditional secondary schools after just a year of secondary school.

A. Regression Analysis

In this section we conduct analysis of telesecundarias using Ordinary Least Squares (OLS) and Instrumental Variables (IV) regressions that use relative distance as an IV for telesecundaria attendance. Column (1) of Table 2 presents the results of a regression of seventh grade math ENLACE exam scores on an indicator for telesecundaria attendance. The regression shows that telesecundaria students score 3.2 points lower on the exam. Column (2) includes controls for the prior year's math and Spanish ENLACE scores, and shows that, conditional on these measures of cognitive skill, telesecundaria students score 25 points better on the exam than students at traditional schools. The inclusion of a rich set of controls for family background and geographic location, in Column (3), does not significantly affect the coefficient on the telesecundaria dummy. Columns (4) - (6) instead show the results of IV regressions of seventh grade math scores on telesecundaria attendance with the full set of controls from column (3) and instruments for telesecundaria attendance. Column (4) uses a single IV that is linear in relative distance, while columns (5) and (6) add quadratic and cubic functions of distance.[12] In all specifications, the first stage F-statistic is large and Hausman tests reject equivalence of the OLS and IV estimates. The Sargan over-identifying test rejects the null hypothesis in the quadratic and cubic specifications. Under the maintained assumption that the instruments are valid (which we discuss in section IV.D), rejecting the null provides evidence of heterogeneous treatment effects (Parente and Silva, 2012).

Table 3 displays the corresponding OLS and IV regressions for Spanish as an outcome variable. The regression coefficients are smaller for Spanish than for math, but still sizable, between 18 and 20 points for the IV specifications. Unlike for math, the IV regressions for Spanish scores do not always reject the

[11]The conditional cash transfer program in Mexico began in 1997 and has been has been called PROGRESA, Oportunidades, and Prospera. The main educational component of the program is a cash transfer that families receive if their child is enrolled in school. Parker and Todd (2017) provide a review of the literature on the effects of conditional cash transfer in Mexico and conclude that it has been effective in increasing school enrollment, reducing grade retention, and increasing educational attainment.

[12]Since relative distance can be both positive and negative, we compute our quadratic instrument by $dist^2_{tele} - dist^2_{trad}$ to preserve the sign of the instrument, and similarly our cubic instrument is $dist^3_{tele} - dist^3_{trad}$.

Hausman test at the 5% significance level. As a result, when we explore sorting into telesecundarias on the basis of observables in later sections, we will primarily analyze the math outcome variable.

Overall, the OLS and IV regressions that control for lagged test scores in Tables 2 and 3 suggest, like Figure 1, that students at telesecundarias experience greater learning growth than students in traditional schools. The Sargan and Hausman test results additionally provide evidence of heterogeneous treatment effects and sorting based on unobserved gains, echoing prior findings in the literature evaluating schooling transitions (Heckman and Vytlacil, 2007). In the next section, we outline an empirical framework that formally models these two features.

TABLE 2—OLS/IV REGRESSIONS: MATH

	Dependent variable: Math Score (Grade 7)					
	OLS	OLS	OLS	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Telesecundaria	−3.22	25.0	26.2	30.4	31.6	30.0
	(0.759)	(0.600)	(0.705)	(1.48)	(1.35)	(1.30)
Math Score (6th Grade)		0.425	0.415	0.415	0.415	0.415
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Spanish Score (6th Grade)		0.222	0.219	0.219	0.219	0.219
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Controls	No	No	Yes	Yes	Yes	Yes
Instrument				Linear	Quadratic	Cubic
First Stage F Statistic				34853	20957	15360
(p value)				(0.000)	(0.000)	(0.000)
Hausman Test				11.9	24.2	13.3
(p value)				(0.001)	(0.000)	(0.000)
Over-identification					6.25	29.0
(p value)					(0.012)	(0.000)

Note: The table shows OLS and IV regressions with Grade 7 math scores as the outcome variable. The instrument is relative distance, and is incorporated using linear, quadratic and cubic functions. Columns (3) -(6) include the full set of controls: average test scores at the nearest telesecundaria and traditional school, the student's age, gender, number of siblings, conditional cash transfer status, family income level, mother's education, indigenous status, number of books at home, computer at home, a rural dummy, a northern state dummy, and five dummies for municipality population. Standard errors are robust to heteroskedasticity. First stage regressions for columns (4), (5), and (6) are shown in Table B-1 in Appendix B.

TABLE 3—OLS/IV REGRESSIONS: SPANISH

	Dependent variable: Spanish Score (Grade 7)					
	OLS	OLS	OLS	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Telesecundaria	−13.4	15.5	16.4	18.7	19.3	18.2
	(0.744)	(0.590)	(0.708)	(1.47)	(1.36)	(1.31)
Math Score (6th Grade)		0.111	0.121	0.121	0.121	0.121
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Spanish Score (6th Grade)		0.531	0.484	0.484	0.484	0.484
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Controls	No	No	Yes	Yes	Yes	Yes
Instrument				Linear	Quadratic	Cubic
First Stage F Statistic				34853	20957	15360
(p value)				(0.000)	(0.000)	(0.000)
Hausman Test				3.53	6.83	2.85
(p value)				(0.060)	(0.009)	(0.091)
Over-identification					1.51	12.6
(p value)					(0.219)	(0.002)

Note: The table shows OLS and IV regressions with Grade 7 Spanish scores as the outcome variable. The instrument is relative distance, and is incorporated using linear, quadratic and cubic functions. Columns (3) -(6) include the full set of controls: average test scores at the nearest telesecundaria and traditional school, the student's age, gender, number of siblings, conditional cash transfer status, family income level, mother's education, indigenous status, number of books at home, computer at home, a rural dummy, a northern state dummy, and five dummies for municipality population. Standard errors are robust to heteroskedasticity. First stage regressions for columns (4), (5), and (6) are shown in Table B-1 in Appendix B.

III. Model

A. Student Achievement

We apply the potential outcomes framework of Rubin (1974) to a value-added model of learning. Students can either attend a telesecundaria or a traditional school.[13] We define the random variable D, where $D = 1$ denotes attendance in a telesecundaria and $D = 0$ denotes attendance in a traditional school. We study the effects of telesecundaria attendance on two outcomes: math and Spanish test scores in seventh grade. For each course $C \in \{\text{Math, Spanish}\}$, the potential outcomes Y_0^C and Y_1^C correspond to the grade seven test score a student would achieve had she enrolled in a traditional or telesecundaria school, respectively. For ease of notation, we will omit the C superscripts. The same causal model will be used for each of the two outcomes.

We model test score outcomes and school choice according to the selection model in Heckman and Vytlacil (2005):

$$Y_1 = X\beta_1 + U_1 \ , \tag{1}$$

$$Y_0 = X\beta_0 + U_0 \ , \tag{2}$$

$$D = \mathbb{1}(\mathbf{Z}\gamma > V) \ , \tag{3}$$

where X is a vector of observable characteristics influencing outcomes, $\mathbf{Z}$ is a vector of observable characteristics influencing the choice of secondary school, and (U_1, U_0, V) are unobserved by the econometrician. Students may have partial information that is not available to the econometrician on the future realizations of (U_1, U_0, V) and may act upon this information. Equations (1) and (2) are value-added equations: X contains the previous year's test scores in both math and Spanish.[14]

The effect of attending a telesecundaria relative to a traditional school on a student's test scores is given by $Y_1 - Y_0$, and the average effect for individuals with a specific set of observable characteristics is $ATE(X) = \mathbb{E}[Y_1 - Y_0 | X = x]$. The fundamental challenge in estimating any sort of treatment effect is that the econometrician only observes one of the two potential outcomes, $Y = Y_0 + D(Y_1 - Y_0)$.

[13]Mexico has three public secondary school types: general, technical, and telesecundaria. We consider the choice between a traditional school (general/technical) and a telesecundaria. Table B-2 in Appendix B reveals that general and technical schools have similar distributions of observable household characteristics and student test scores in the seventh grade, so we believe that it is reasonable to consider them as a single alternative for the purposes of evaluating learning in math and Spanish between the sixth and seventh grades. However, the ensuing analysis goes through without modification if general and technical schools are treated as separate alternatives as long as the results are re-interpreted as the causal effect of telesecundaria education relative to the next best alternative.

[14]The value-added framework admits a causal interpretation when the effects of time-varying investments on test scores decline geometrically with the time between when the investment was made and when the test was taken (Boardman and Murnane, 1979; Todd and Wolpin, 2003).

The instrument, Z, is assumed to satisfy an exclusion restriction, meaning that it must simultaneously induce variation in the choice of school conditional on the covariates, X, and have no direct effect on the outcome variables. When Z is a sufficiently predictive instrument, as in this paper, it can shift the probability of attending a telesecundaria continuously between 0 and 1. Such a shift would cause even the most unlikely student to attend a telesecundaria, thereby permitting a comparison of Y_0 and Y_1 for all students.

The marginal treatment effect (MTE), introduced by Björklund and Moffitt (1987) and extended in Heckman and Vytlacil (1999; 2001*b*; 2005; 2007), is the average treatment effect for an individual at a particular margin of "resistance to treatment." V, in equation (3), represents this resistance to treatment. An individual with a higher V is, on the basis of unobservables, less likely to attend a telesecundaria. As in Heckman and Vytlacil (2005), we apply the following useful transformation to equation (3) to obtain

$$\begin{aligned} D &= \mathbb{1}(Z\gamma > V) \\ &= \mathbb{1}(F_V(Z\gamma) > F_V(V)) \\ &= \mathbb{1}(P(Z) > U_D) \ , \end{aligned}$$

where $P(Z)$ is the propensity score and $U_D \sim U[0,1]$. Following the transformation, the MTE can be written as

$$MTE(x, u_D) = \mathbb{E}[Y_1 - Y_0 | X = x, U_D = u_D] \ . \tag{4}$$

The marginal treatment effect measures the average difference in outcomes at telesecundarias relative to traditional schools for individuals with observable characteristics x and latent resistance to treatment u_D.

B. Identification

The MTE is identified under the following assumptions, stated in Heckman and Vytlacil (2005) and modified slightly to fit this paper's notation:

- **(A-1)** Z is a nondegenerate random variable conditional on X.
- **(A-2)** $(U_0, U_1, U_D) \perp\!\!\!\perp Z | X$.
- **(A-3)** U_D is absolutely continuous with respect to the Lebesque measure.
- **(A-4)** $\mathbb{E}|Y_1|$ and $\mathbb{E}|Y_0|$ are finite.
- **(A-5)** $1 > \mathbb{P}(D = 1|X) > 0$.

Assumption (A-1) ensures that the instrument influences attendance in telesecundarias conditional on covariates, X, while (A-2) assumes that the instrument is exogenous in the sense that it is independent of unobservable variables in the selection and outcome equations conditional on X. Assumptions (A-3) - (A-5) are technical assumptions that are satisfied in our setting. Under these assumptions, $MTE(X, U_D)$ is nonparametrically identified by the Local Instrumental Variables (LIV) estimand (Heckman and Vytlacil, 2001*a*):

$$\frac{\partial \mathbb{E}[Y|X=x, P(z)=p]}{\partial p}\Big|_{p=u_D} = MTE(x, u_D) \; . \tag{5}$$

The LIV estimand shows that the marginal treatment effect at each value of the latent resistance to treatment, U_D, is identified by individuals who are indifferent between being treated and not, because when $p = u_D$, the individual is on the knife edge between participating and not participating.

C. *Parameters of Interest*

A large class of parameters corresponding to the effect of telesecundaria attendance on schooling outcomes can be written as weighted averages of the marginal treatment effect. In this paper we are interested in $ATE(X)$, as well as the average effect of treatment on the treated, $TT(X) = \mathbb{E}[Y_1 - Y_0 \mid X = x, D = 1]$, the average effect of treatment on the untreated, $TUT(X) = \mathbb{E}[Y_1 - Y_0 \mid X = x, D = 0]$, and various treatment effects corresponding to the effects of never-before implemented policies. These Policy-Relevant Treatment Effects, first defined in Heckman and Vytlacil (2001*b*), are defined for a shift from a pre-existing policy a to a new policy a' and provide a normalized effect of the policy change:

$$PRTE_{a',a}(X) = \frac{\mathbb{E}[Y_{a'} - Y_a \mid X = x]}{\mathbb{P}(D_{a'} = 1 \mid X = x) - \mathbb{P}(D_a = 1 \mid X = x)} \; . \tag{6}$$

Any treatment parameter, including ATE, TT, TUT, and $PRTE_{a',a}$ for policies a and a', can be computed by integrating the MTE with respect to the distribution of U_D induced by the treatment parameter under consideration:

$$ATE = \int MTE(X, U_D) dF(X, U_D) \; ,$$

$$TT = \int MTE(X, U_D) dF_{U_D, X|D=1}(x, u_D \mid D = 1) \; , \tag{7}$$

$$TUT = \int MTE(X, U_D) dF_{U_D, X|D=0}(x, u_D \mid D = 0) \; , \tag{8}$$

$$PRTE_{a',a} = \int MTE(X, U_D) dF_{U_D, X|D_a=0, D_{a'}=1}(x, u_D \mid D_a = 0, D_{a'} = 1) \; . \tag{9}$$

In the next section, we discuss the methods we use to integrate the MTE to obtain these treatment parameters.

D. Estimation

Assumptions (A-1) - (A-5) require that $MTE(X, U_D)$ be estimated separately for each X. When X is high-dimensional, as in our setting, $MTE(X, U_D)$ is only identified by the support of $P(Z)$ given X. Even if the unconditional support of $P(Z)$ is the entire unit interval, $supp(P(Z)) \mid X = x)$ may consist of only a few points. Since this will be too few to estimate $MTE(X, U_D)$ with any degree of precision, we strengthen assumption (A-2) to (A-2)′:

(A-2)′ $(X, Z) \perp\!\!\!\perp (U_1, U_0, U_D)$

Assumption (A-2)′ is standard in the literature estimating selection models. It has two consequences. The first is that, together with the linear framework in equations (1) and (2), it yields an MTE function that is additively separable in X and U_D so that

$$MTE(X, U_D) = X(\beta_1 - \beta_0) + \kappa(U_D) \ .$$

An additional consequence of assumption (A-2)′ is that $MTE(X, U_D)$ can now be identified on the unconditional support of $P(Z)$ rather than $supp(P(Z) \mid X)$. The cost of the assumption is that it restricts the pattern of selection on unobservables – given by the shape of $MTE(X, \cdot)$ – to be the same across individuals with different observable characteristics, X. It rules out the possibility that $MTE(X, \cdot)$ has a different slope depending on the value of X (level shifts can be accommodated). Sensitivity analysis in Appendix D conditions on subsamples defined by different variables in X. Reassuringly, the shape of sorting on gains does not vary much across the different subsamples, suggesting that assumption (A-2)′ is not overly strong in our setting.

We estimate the MTE using two methods: a parametric approach that specifies the joint distribution of unobservables and a semiparametric approach that leaves the joint distribution of unobservables unspecified and estimates $\frac{\partial \mathbb{E}[Y|X=x,P(z)=p]}{\partial p}$ using local polynomial modeling. We estimate the parametric approach via a two-step "Heckit" procedure: We refer the reader to Appendix C for details. The semiparametric approach, the LIV estimator of Heckman and Vytlacil (2001*a*), involves estimating $\frac{\partial \mathbb{E}[Y|X=x,P(z)=p]}{\partial p}$ using the partially linear model estimator of Robinson (1988). To understand this method note that assumptions (A-1), (A-2)′, (A-3) - (A-5), together with the assumption that the outcome models in (1) and (2) are linear, yields a conditional expectation function for Y that is linear in X and XP and nonlinear in the propensity score, P:

$$\begin{aligned}\mathbb{E}[Y|X=x,P(z)=p] =&\mathbb{E}[Y_0+D(Y_1-Y_0)|X=x,P(z)=p]\\ =&X\beta_0+\mathbb{E}[DX(\beta_1-\beta_0)|X=x,P(z)=p]+\\ &\mathbb{E}[U_0+D(U_1-U_0)|X=x,P(z)=p]\\ =&X\beta_0+PX(\beta_1-\beta_0)+K(P)\ ,\end{aligned}$$

where $K(P)$ is a nonparametric function of the propensity score. This form for the conditional expectation means that the marginal treatment effect is given by

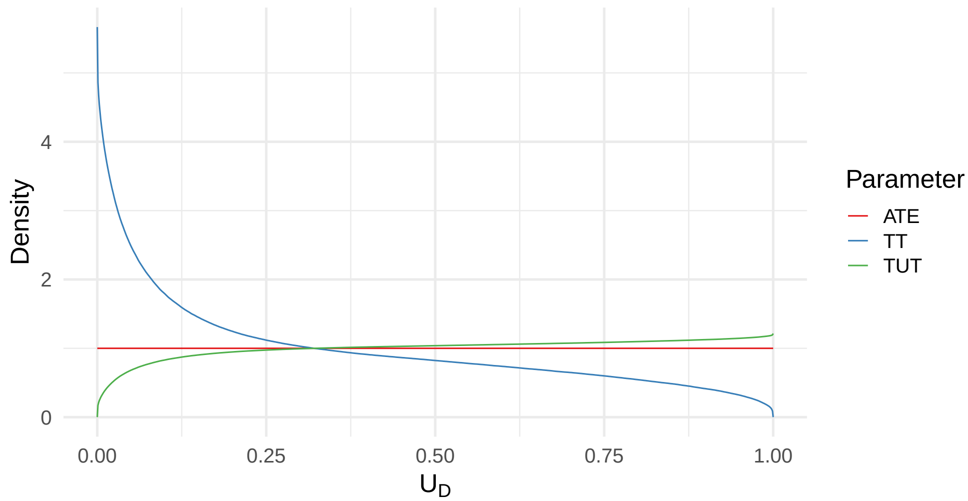
$$MTE(X,U_D)=X(\beta_1-\beta_0)+\frac{\partial K(P)}{\partial P}\Big|_{p=U_D}\ . \tag{10}$$

The Robinson (1988) semiparametric estimator of (10) entails two steps. First, the estimated propensity score, P, is partialed out of the other variables by running nonparametric regressions of Y, X, and PX on P. Then the residualized Y is regressed linearly on the residualized X and PX to obtain estimates of β_0 and $\beta_1-\beta_0$. In the second step, the derivative of the conditional expectation of $\hat{Y}\equiv Y-X\hat{\beta}_0-XP(\hat{\beta}_1-\hat{\beta}_0)$ with respect of P is estimated nonparametrically to obtain an estimate of $\frac{\partial K(P)}{\partial P}$.

All nonparametric regressions are estimated using local polynomial regression. Following the recommendations in Fan and Gijbels (1996), we use local linear regression to estimate the conditional expectations in the first stage and local quadratic regression to estimate $\frac{\partial K(P)}{\partial P}$ in the second stage. We use the plug-in estimator of Fan and Gijbels (1996) to select the bandwidth. This bandwidth, which aims to minimize the Integrated Mean Square Error (IMSE) in the final nonparametric regression, depends negatively on the function's second derivative and on the density of the data, and positively on the conditional variance of the outcome variable. The plug-in method selects a bandwidth of 0.231 for math test scores and 0.323 for Spanish scores.

We estimate treatment parameters by integrating the semiparametric $MTE(X,U_D)$ with respect to the appropriate distributions in equations (7) - (9) using the simulation method introduced in Carneiro, Lokshin and Umapathi (2017). The simulation approach, which is only valid under assumption (A-2)′, involves creating an equally-spaced grid for U_D for each individual and averaging $MTE(X,U_D)$ for the values of U_D on the grid that are less than that individual's propensity score, $P(Z)$, for TT, greater than that individual's propensity score for TUT, and between $P(Z_a)$ and $P(Z_{a'})$ for $PRTE_{a',a}$. Figure 3 displays the densities used to compute ATE, TT, and TUT, plotted as a function of U_D. The figure shows that ATE uniformly samples individuals with all levels of U_D, while TT oversamples individuals with low U_D, and TUT oversamples individuals with high U_D.

FIGURE 3. TREATMENT PARAMETER WEIGHTS



Note: The figure shows the distribution of weighting functions used to construct three standard treatment parameters. The average treatment effect (ATE) integrates the MTE with respect to the unit uniform distribution. The average effect of treatment on the treated (TT) integrates the MTE with respect to the distribution of U_D conditional on attendance in telesecundarias, $f_{U_D,X|D=1}(x, u_D \mid D = 1)$, while the average effect of treatment on the untreated (TUT) integrates the MTE with respect to the distribution of U_D conditional on attendance in traditional schools, $f_{U_D,X|D=0}(x, u_D \mid D = 0)$.

E. Is Relative Distance a Valid Instrument?

Identification of the marginal treatment effect requires that the instrument satisfy assumptions (A-1) and (A-2)′. The first assumption, that relative distance predict attendance in telesecundarias conditional on observable covariates, X, is easily verified. Table 4 displays the average marginal effects of each variable in the propensity score model on the probability of attending a telesecundaria (estimated via probit). The average marginal effect of relative distance is a statistically significant increase in the probability of attendance of 2.9 percentage points per kilometer.

The second assumption, that the instrument be independent of unobservable variables in the outcome and selection equations (U_1, U_0, V), cannot be proven directly. Our main outcome is test scores, so it would be concerning if the instrument were correlated with measures of student ability or school quality, as these would likely affect achievement. Table 5 reports the results of balancing tests that examine whether the instrument is correlated with observed measures of student ability. The first row of the table presents the estimates of univariate regressions of sixth grade math and Spanish scores on the instrument. The coefficients are statistically significant, but not economically large, with each km increase in rela-

TABLE 4—PROPENSITY SCORE MODEL

	Average Derivative	Standard Error
Relative Distance	-0.029	(0.000)
Math Score (6th Grade)	-0.006	(0.001)
Spanish Score (6th Grade)	-0.010	(0.001)
Mean Score: nearest Telesecundaria	0.013	(0.002)
Mean Score: nearest Traditional	0.007	(0.002)
Age	0.015	(0.000)
Siblings	0.004	(0.000)
Female	0.003	(0.001)
Prospera	0.024	(0.002)
Computer	-0.022	(0.002)
Rural Residence	0.021	(0.002)
Northern State	-0.029	(0.003)
Speaks Language other than Spanish at Home	0.001	(0.003)
Family Income : Low	-0.012	(0.002)
Family Income : Medium	-0.016	(0.002)
Family Income : High	-0.022	(0.003)
Mother's Education : Middle	-0.016	(0.002)
Mother's Education : Secondary	-0.034	(0.003)
Mother's Education : Post-Secondary	-0.017	(0.005)
Books in the Home : 20	-0.011	(0.002)
Books in the Home : 50	-0.018	(0.003)
Books in the Home : $\geq$100	-0.012	(0.003)
Municipality : 10K-20K	-0.000	(0.005)
Municipality : 20K-50K	0.001	(0.004)
Municipality : 50K-100K	0.002	(0.004)
Municipality : 100K-500K	-0.016	(0.004)
Municipality : $>$ 500K	-0.026	(0.004)

Note: The table shows the average marginal effects of each variable in the propensity score model for telesecundaria attendance. The instrument, relative distance, is computed as the difference between two distance measures. The first is the distance from the student's primary school to the nearest telesecundaria, and the second is the distance from the student's primary school to the nearest traditional school. Relative distance is negative whenever telesecundarias are closer. Test score variables refer to the effects of increases of 100 points (about 1 sd). The omitted category in each of Family Income, Mother's Education, Books in the Home, and Municipality is the lowest one. Computer is a binary variable that equals one if the student has access to a computer at home. Standard errors are calculated via 250 bootstrap replications.

tive distance associated with a 1.6-1.8 point increase in test scores. However, after including the controls in our propensity score model, the relationship between relative distance and sixth grade test scores becomes negligible and statistically insignificant.

TABLE 5—RELATIONSHIP BETWEEN THE INSTRUMENT AND LAGGED TEST SCORES

Controls	Math (1)	Spanish (2)
None	1.63	1.82
	(0.075)	(0.066)
Model Covariates	0.114	−0.015
	(0.094)	(0.079)
Observations	113,525	

Note: The table reports coefficients from regressions of grade 6 math and Spanish test scores on the instrument, relative distance. The first row does not control for any covariates, while the second row includes all covariates from the propensity score model.

In Appendix G, we additionally show that, after controlling for municipality population and a rural dummy, relative distance is uncorrelated with the quality of the nearest school, where we measure school quality by the average 8th grade ENLACE score for an older cohort. These balancing tests lend plausibility to the assumption that, after controlling for the rich set of observable characteristics in our analysis, relative distance is independent of unobserved determinants of test score outcomes.

An extensive literature exploits geographic instruments like distance to evaluate Catholic high schools in the United States (Neal, 1997; Grogger and Neal, 2000). Altonji, Elder and Taber (2005) show how geographic instruments strongly predict high school graduation in a sample of students who are very unlikely to attend Catholic school, suggesting that these instruments may be correlated with unobserved determinants of schooling outcomes. We do not find evidence of this in our context. Table 6 examines the relationship between relative distance and three secondary school outcomes in a sample of students who are very unlikely to attend telesecundarias (those for whom the nearest telesecundaria is over 15 km away). Out of $2,139$ such students, only ten attend telesecundarias ($< 0.5\%$). The table shows that there is no discernible relationship between relative distance and secondary school outcomes for these students and provides further evidence of the validity of the exclusion restriction.

The MTE can only be nonparametrically identified where the propensity score has positive support for both treated and untreated individuals. The region of common support in this paper, depicted in Figure 4, is the full $[0, 1]$ interval, owing to both the large sample size and the strong instrument. Each ventile of the propensity score's distribution contains over 150 treated and control individuals, permitting relatively precise estimation of the MTE at both the interior and boundaries of its support.

TABLE 6—EFFECT OF RELATIVE DISTANCE ON SECONDARY SCHOOL OUTCOMES

	Math Score (7th Grade)			*Spanish Score (7th Grade)*			*Grade 9 Enrollment*		
Relative Distance	0.545	0.504	1.02	0.877	0.376	0.179	0.003	0.002	0.000
	(0.835)	(0.638)	(0.721)	(0.827)	(0.658)	(0.732)	(0.003)	(0.003)	(0.003)
Math Score (6th Grade)		0.396	0.385		0.068	0.087		0.000	0.000
		(0.024)	(0.025)		(0.022)	(0.023)		(0.000)	(0.000)
Spanish Score (6th Grade)		0.227	0.233		0.575	0.508		0.000	0.000
		(0.026)	(0.029)		(0.025)	(0.028)		(0.000)	(0.000)
Controls	No	No	Yes	No	No	Yes	No	No	Yes

Note: The table shows OLS regressions of three secondary school outcomes on the instrument, relative distance, for the 2,139 students whose nearest telesecundaria is over 15 km away. Controls include sex, number of siblings, mother's education, family income, number of books in the home, family access to a computer, Prospera status, rural residence, residence in a Northern state, whether the family speaks Spanish at home, a measure of school quality, and dummies for municipality size. Standard errors are robust to heteroskedasticity.

IV. Empirical Results

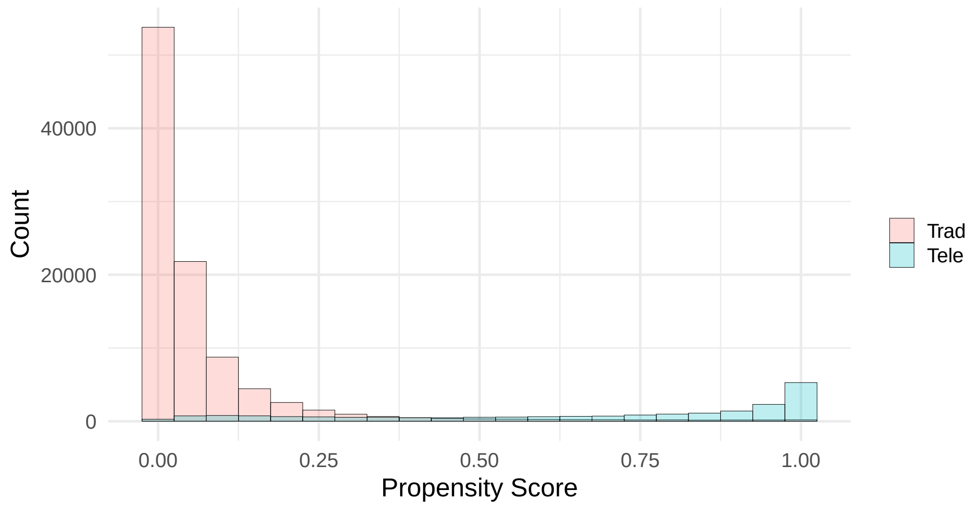
A. Main Findings

This section presents the results of the MTE analysis. Figure 5 shows the estimated MTEs for seventh grade math scores evaluated at mean values of X. The parametric and semiparametric MTEs are plotted side-by-side with 90% confidence bands in grey. Figure 6 repeats the analysis with Spanish as the outcome variable. In both figures, the horizontal axis measures the latent variable, U_D, while the vertical axis measures the expected benefit to attending a telesecundaria relative to a traditional school for students with that level of U_D, $\mathbb{E}[Y_1 - Y_0|X = \overline{x}, U_D = u_D]$.

The MTE curves reveal a pattern of negative sorting on gains. Students who, on the basis of unobservables, are most likely to attend telesecundarias (low U_D) experience little value added in both math and Spanish. As U_D increases, the benefits of telescundaria attendance increase sharply. For large values of U_D, the math semiparametric MTE is mostly flat, while the Spanish MTE has large, but statistically insignificant, increases as U_D approaches 1. Both MTEs show that the treatment effects are nonnegative for nearly all values of U_D.

Tables 7 and 8 present estimates of standard treatment parameters for math and Spanish, respectively. All treatment parameters are positive, underscoring the findings from the MTE curve that telesecundaria attendance is beneficial for a large majority of students. Standard errors reveal that the treatment parameters are precisely estimated and are significant at conventional levels of significance.

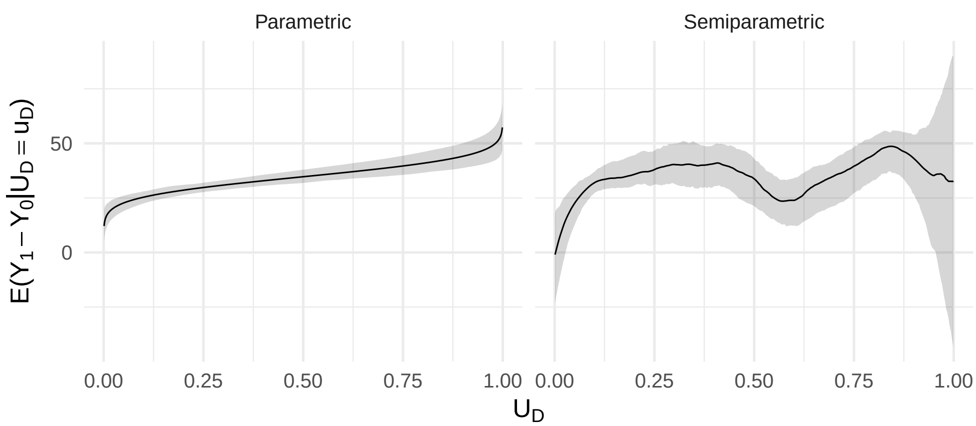
FIGURE 4. ESTIMATED PROPENSITY SCORE BY SECONDARY SCHOOL ATTENDED



Note: The figure plots histograms of estimated propensity scores by the type of secondary school attended. The propensity score models telesecundaria attendance as a function the child's sixth grade math and Spanish scores, age, sex, number of siblings, mother's education, family income, number of books in the home, family access to a computer, Prospera status, rural residence, residence in a Northern state, whether the family speaks Spanish at home, a measure of school quality, dummies for municipality size, and the relative distance between the nearest telesecundaria and nearest traditional school. The propensity score model is estimated via probit. The common support of propensity scores across treated and control units is the full $[0, 1]$ interval.

The semiparametric estimate of the ATE for math indicates that a randomly selected student would be expected to perform 35.0 points better in mathematics in the seventh grade had she attended a telesecundaria instead of a traditional school. As the standard deviation on the 7th grade math ENLACE exam is 100.7, this represents an increase of 0.347 standard deviations. The estimates of TT (27.4 points) are smaller than those of TUT (36.6 points), reinforcing the finding of negative sorting on gains for math. These numbers indicate that the average student currently attending a telesecundaria gains 0.272 standard deviations while the average student currently attending a traditional school would gain 0.363 standard deviations on the math ENLACE exam if she were to attend a telesecundaria. The semiparametric estimates of ATE, TT, and TUT for Spanish are 22.5, 19.0, and 23.2 points, which represent 0.225, 0.190, and 0.232 standard deviations, respectively. These estimates, while smaller than those for math, are significantly different from zero. In addition, while the Spanish MTE is nonmonotonic, the finding that TT$<$ATE$<$TUT indicates that the pattern of sorting on gains is primarily negative.

FIGURE 5. MARGINAL TREATMENT EFFECT: MATH



Note: The dependent variable in the outcome equation is the raw score on the seventh grade ENLACE math exam. The outcome equations include controls for sixth grade math and Spanish scores, age, sex, number of siblings, mother's education, family income, number of books in the home, family access to a computer, Prospera status, rural residence, residence in a Northern state, whether the family speaks Spanish at home, a measure of school quality, and dummies for municipality size. The school choice model includes the same controls and also includes the relative distance between the nearest telesecundaria and nearest traditional secondary school as an instrument. The school choice model is estimated via probit. The parametric MTE is estimated using a two-step "Heckit" procedure. The semiparametric MTE is estimated using Local Quadratic Regression and an Epanechnikov kernel with a bandwidth of 0.231. Both MTEs are evaluated at the mean value of the covariates, $X = \overline{x}$. Ninety percent confidence intervals are computed via nonparametric bootstrap with 250 draws.

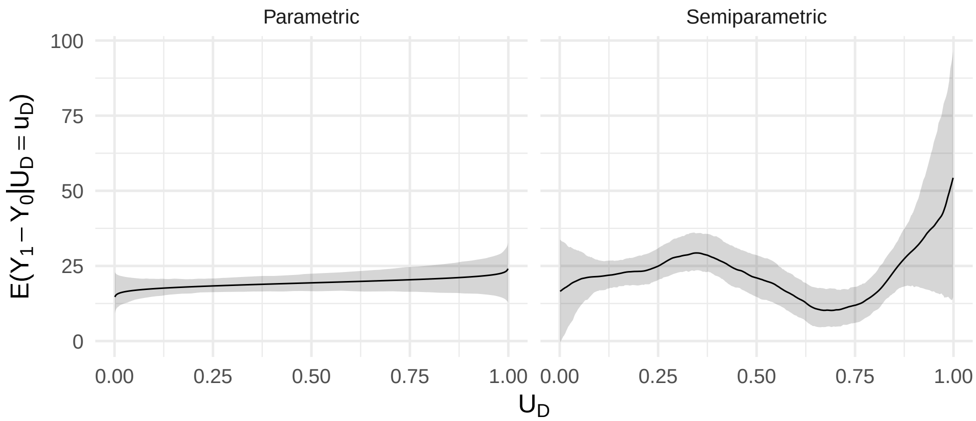
B. Explanations for Sorting on Gains

Figures 5 and 6 reveal a pattern of reverse sorting on gains, which occurs when students whose unobservables, U_D, make them most likely to attend telesecundarias receive the smallest benefits from them. Such a pattern cannot be explained by a choice model in which individuals attend telesecundarias if $Y_1 > Y_0$, as in the original Roy (1951) model. Brinch, Mogstad and Wiswall (2017) show how it is possible to decompose the marginal treatment effect curve into two separate functions representing sorting only into telesecundarias and sorting only into traditional schools, respectively. The authors note that the MTE can be rewritten as

$$MTE(X, U_D) = X(\beta_1 - \beta_0) + k_1(U_D) - k_0(U_D)\ , \tag{11}$$

where $k_j(U_D) = \mathbb{E}[U_j|U_D]$ for $j = 1, 2$. Here $k_1(U_D)$ can be thought of as the average unobserved match quality between students and telesecundarias for

FIGURE 6. MARGINAL TREATMENT EFFECT: SPANISH



Note: The dependent variable in the outcome equation is the raw score on the seventh grade ENLACE Spanish exam. The outcome equations include controls for sixth grade math and Spanish scores, age, sex, number of siblings, mother's education, family income, number of books in the home, family access to a computer, Prospera status, rural residence, residence in a Northern state, whether the family speaks Spanish at home, a measure of school quality, and dummies for municipality size. The school choice model includes the same controls and also includes the relative distance between the nearest telesecundaria and nearest traditional secondary school as an instrument. The school choice model is estimated via probit. The parametric MTE is estimated using a two-step "Heckit" procedure. The semiparametric MTE is estimated using Local Quadratic Regression and an Epanechnikov kernel with a bandwidth of 0.323. Both MTEs are evaluated at the mean value of the covariates, $X = \overline{x}$. Ninety percent confidence intervals are computed via nonparametric bootstrap with 250 draws.

students with a particular resistance to attending telesecundarias (given by U_D). Similarly, $k_0(U_D)$ represents the unobserved match quality between students and traditional schools as a function of the same unobserved variable.

We estimate $k_1(U_D)$ and $k_0(U_D)$ separately to determine whether the shape of the MTE curve is determined primarily by variability in match qualities between students and telesecundarias or between students and traditional schools. As originally noted in Heckman and Vytlacil (2007), $k_1(U_D)$ and $k_0(U_D)$ can be estimated using a control function approach on each of the $D = 1$ and $D = 0$ subsamples. Under Assumption (A-2)′,

$$\mathbb{E}[Y_j|X = x, P(Z) = p, D = j] = X\beta_j + K_j(p) \ ,$$

for $j = 0, 1$, where

$$K_1(p) = E\left(U_1 \mid U_D \leq p\right) \ ,$$

TABLE 7—ESTIMATED TREATMENT EFFECTS: MATH

	Parametric		Semiparametric	
	Estimate	Standard Error	Estimate	Standard Error
Average Treatment Effect	34.7	(1.80)	35.0	(2.68)
Treatment on the Treated	30.4	(1.45)	27.4	(2.20)
Treatment on the Untreated	35.7	(1.93)	36.6	(3.08)

Note: The table displays three treatment parameters corresponding to the effect of telesecundaria attendance on raw scores on the seventh grade math ENLACE exam. The three treatment parameters are obtained by integrating the MTE with respect to the densities displayed in Figure 3. The simulation method of Carneiro, Lokshin and Umapathi (2017) is used to integrate the semiparametric MTE. Standard errors are obtained by 250 bootstrap replications.

TABLE 8—ESTIMATED TREATMENT EFFECTS: SPANISH

	Parametric		Semiparametric	
	Estimate	Standard Error	Estimate	Standard Error
Average Treatment Effect	19.4	(1.78)	22.5	(2.71)
Treatment on the Treated	18.5	(1.50)	19.0	(2.10)
Treatment on the Untreated	19.6	(1.90)	23.2	(3.06)

Note: The table displays three treatment parameters corresponding to the effect of telesecundaria attendance on raw scores on the seventh grade Spanish ENLACE exam. The three treatment parameters are obtained by integrating the MTE with respect to the densities displayed in Figure 3. The simulation method of Carneiro, Lokshin and Umapathi (2017) is used to integrate the semiparametric MTE. Standard errors are obtained by 250 bootstrap replications.

and

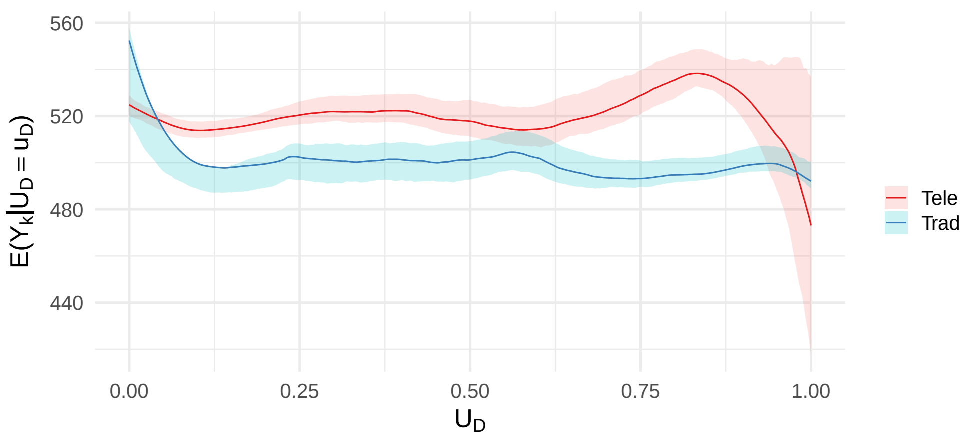
$$K_0(p) = E\left(U_0 \mid U_D > p\right) \ .$$

We can obtain k_1 and k_0 from K_1 and K_0 using the following identities in Brinch, Mogstad and Wiswall (2017):

$$\begin{aligned} k_1(p) &= p\frac{\partial K_1(p)}{\partial p} + K_1(p) \\ k_0(p) &= -(1-p)\frac{\partial K_0(p)}{\partial p} + K_0(p) \ . \end{aligned} \tag{12}$$

Figure 7 shows estimates of k_1, labeled *Tele*, and k_0, labeled *Trad*, for math as an outcome variable. Each estimated curve, k_j for $j = 0, 1$, is obtained via two semiparametric regressions on the subsample with $D = j$. We compute the conditional expectation of $Y - X\beta_j$ given the estimated propensity score, p, using

FIGURE 7. DECOMPOSING THE SOURCES OF REVERSE SORTING ON GAINS



Note: The figure plots $k_1(U_D) = \mathbb{E}[U_1 \mid U_D]$, labeled Tele, and $k_0(U_D) = \mathbb{E}[U_0 \mid U_D]$, labeled Trad, evaluated at $X = \overline{x}$ on the vertical axis against U_D on the horizontal axis. U_1 is the child's unobserved outcome in the equation for math value added in telesecundarias. U_0 is the child's unobserved outcome in the equation for math value added in traditional schools. Details on the estimation of $k_1(\cdot)$ and $k_0(\cdot)$ are provided in section IV.B

Local Linear Regression to obtain $K_j(p)$ and the derivative of the conditional expectation of $Y - X\beta_j$ given p via Local Quadratic Regression to obtain $\frac{\partial K_j(p)}{\partial p}$. $k_j(p)$ is then obtained from the identities in (12). The bandwidth, 0.231, is the same as in the estimation of the MTE function in the previous section. The 90% confidence intervals, obtained via 250 bootstrap replications, are considerably wider than when estimating the original MTE functions in the previous section, because each curve requires two applications of the Robinson double residual semiparametric regression method on subsamples with fewer observations than the full sample.

Despite the wide confidence intervals, it is evident from Figure 7 that the main variation in the MTE curve in Figure 5 is coming from sorting into traditional schools. The function governing outcomes at telesecundarias, $k_1(U_D)$, is instead flat over nearly the entire domain, apart from a statistically insignificant downward-sloping pattern for high levels of U_D.

This pattern suggests that there is considerably more variation in unobserved outcomes at traditional schools than telesecundarias, a finding that is reasonable given that lectures and learning materials are uniform across telesecundarias. Findings of reverse sorting on gains are not unprecedented in the literature. Cornelissen et al. (2018) and Ainsworth et al. (2022) find evidence of parents choosing academically inferior educational options for their children, often due to a lack of information or inaccurate beliefs. Abdulkadiroğlu et al. (2020) argue that parents do not value school effectiveness conditional on peer quality.

Our findings are consistent with both an information-based explanation and with parents having preferences over factors other than test score gains. Informed parents may be more likely to send their children to telesecundarias (low U_D) while simultaneously being able to ensure that their children have good learning outcomes at both types of schools. This scenario is certainly plausible in Mexico, where most opinions of telesecundarias are negative, but where opinions of a small percentage of informed individuals may be positive.[15] In a separate setting, Walters (2018) argues that reverse sorting on gains into charter schools in Boston may be caused by unobserved parental investment which is more productive if the student attends a traditional school than a charter school, since charter schools weaken the relationship between parental inputs and student achievement.[16]

Figure B-1 in Appendix B shows that there is also reverse sorting on gains into telesecundarias on the basis of observable characteristics. Students who have a high propensity score, indicating a high probability of attending telesecundarias, have observable characteristics that cause them to benefit less from attending them. Table B-4 show estimates of $\hat{\beta}_1 - \hat{\beta}_0$ side-by-side with the marginal effects of each variable on the probability of attending telesecundarias. Reverse sorting on the basis of observables is driven, in part, by mother's education, with children of higher educated parents being less likely to attend telesecundarias but more likely to benefit.

C. Impacts on Enrollment

Our analysis shows that telesecundarias have significant causal effects on student learning between the sixth and seventh grades. One might naturally expect that they also have positive effects on other outcomes that are correlated with test scores. Navarro-Sola (2021) shows that the expansion of telesecundarias caused significant increases in labor market earnings and an increase in formal labor force participation at the expense of employment in the agricultural and informal sectors. While we do not have labor market data, we do follow students through the ninth grade, which is the final year of secondary school and the final

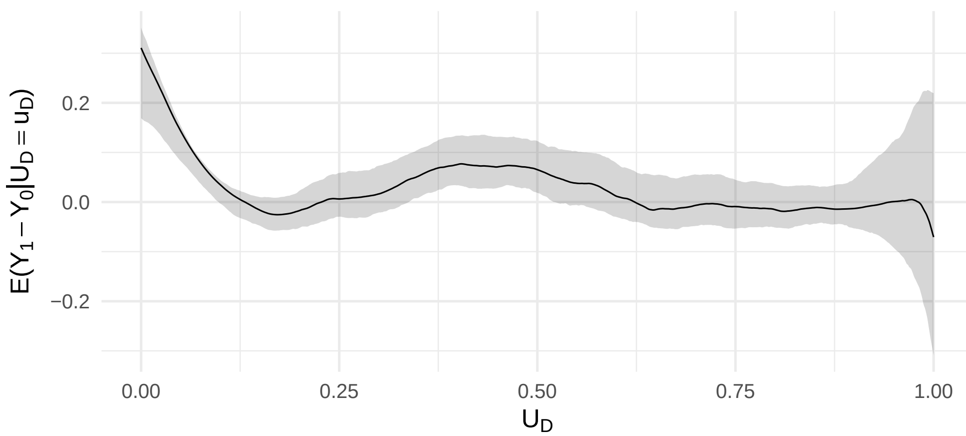
[15]The predominant view expressed in Mexican media and bolstered by academic research is that telesecundarias are of low quality (de Cossio, 2007). Santos (2001) argues that they are worse than traditional schooling options on the grounds that telesecundaria students perform worse academically in the cross section, a finding we confirm for the seventh grade. While this analysis fails to control for the relative deprivation of students who attend telesecundarias, the perception of telesecundarias as lower quality alternatives seems ingrained (Acta Educativa, 2016).

[16]There is some evidence in our data in support of this hypothesis: The parents of children who attend telesecundarias are 6 pp more likely to say that they always read to their child and help their child with homework. However, after including the rich set of controls in our propensity score model, the relationship between school choice and these parental investment variables becomes insignificant, so we decided to omit these variables from the analysis. (Results available from the authors upon request.) The pattern of reverse sorting on gains would therefore need to be caused by harder-to-measure forms of parental investment.

year of compulsory education. Since telesecundarias have positive effects on student learning, it is reasonable to expect that they would also reduce rates of dropout, and this is exactly what we find.

We replicate our main analysis with a dependent variable that equals one if the student attends school in the ninth grade and zero otherwise. Our sample conditions on students who enroll in secondary school, so this analysis answers the following question: Conditional on enrolling in secondary school, does enrollment in telesecundarias cause an increase in the probability of attending school in the ninth grade relative to attendance in traditional secondary schools?

FIGURE 8. MARGINAL TREATMENT EFFECT: NINTH GRADE ENROLLMENT



Note: The dependent variable in the outcome equation is a binary variable that equals one if the individual enrolls in the ninth grade and zero otherwise. The outcome equations include controls for sixth grade math and Spanish scores, age, sex, number of siblings, mother's education, family income, number of books in the home, family access to a computer, Prospera status, rural residence, residence in a Northern state, whether the family speaks Spanish at home, a measure of school quality, and dummies for municipality size. The school choice model includes the same controls and also includes the relative distance between the nearest telesecundaria and nearest traditional secondary school as an instrument. The school choice model is estimated via probit. The MTE is estimated semiparametrically using Local Quadratic Regression and an Epanechnikov kernel with a bandwidth of 0.202. The MTE is evaluated at the mean value of the covariates, $X = \overline{x}$. Ninety percent confidence intervals are computed via nonparametric bootstrap with 250 draws.

We answer this question by estimating a semiparametric MTE and integrating it to compute treatment parameters.[17] Figure 8 displays the estimated MTE for this analysis. The MTE displays a pattern of positive selection on gains: Students who are the most likely to attend telesecundarias, on the basis of having low

[17]The parametric model is misspecified in the case of a binary outcome.

values of U_D, experience larger causal effects on ninth grade enrollment than do students who have a lower likelihood of attending telesecundarias.

TABLE 9—ESTIMATED TREATMENT EFFECTS: NINTH GRADE ENROLLMENT

	Estimate	Standard Error
Average Treatment Effect	0.028	(0.010)
Treatment on the Treated	0.085	(0.008)
Treatment on the Untreated	0.015	(0.011)
Mean : 9th Grade Enrollment	0.915	

Note: The table displays three treatment parameters corresponding to the effect of telesecundaria attendance on ninth grade enrollment. The three treatment parameters are obtained by integrating the semiparametric MTE for ninth grade enrollment with respect to the densities displayed in Figure 3. The simulation method of Carneiro, Lokshin and Umapathi (2017) is used to integrate the MTE. Mean 9th grade enrollment is estimated conditional on attending secondary school for at least one year. Standard errors are obtained by 250 bootstrap replications.

Table 9 presents estimates of ATE, TT, and TUT for the effect of telesecundaria attendance on ninth grade enrollment. These are computed by integrating the semiparametric MTE with respect to the relevant densities in Figure 3. The average treatment effect of telesecundaria attendance on ninth grade completion is 2.8 pp. The average effect is larger for the treated than the untreated, 8.5 versus 1.5 pp. The mean of the dependent variable, which represents the probability of enrolling in the ninth grade conditional on attending secondary school, is 0.915. Telesecundarias therefore have a significant causal effect on reducing dropout during secondary school, which is consistent with them generating learning gains.

D. Functional Form Tests

Assumption (A-2)′ implies that the MTE function is additively separable in X and U_D. This means that the pattern of reverse sorting on gains that we find does not vary with observable characteristics. Kline and Walters (2016) outline a way to test this assumption in a parametric setting. After estimating $\hat{\beta}_0$ and $\hat{\beta}_1$ via the parametric model, we obtain the following residuals for the outcome equations:

$$\hat{U}_1 = Y_1 - X\hat{\beta}_1 \ ,$$
$$\hat{U}_0 = Y_0 - X\hat{\beta}_0 \ ,$$

and regress these residuals on X and interactions between X and selection-correction (inverse mills ratio) terms as follows,

$$\hat{U}_1 = X'\gamma_{11} + \hat{\lambda}_1(P(Z))\gamma_{12} + \hat{\lambda}_1(P(Z))X'\gamma_{12} + \eta_1 \tag{13}$$

$$\hat{U}_0 = X'\gamma_{01} + \hat{\lambda}_0(P(Z))\gamma_{02} + \hat{\lambda}_0(P(Z))X'\gamma_{02} + \eta_0 \ , \tag{14}$$

where $\hat{\lambda}_1(P(Z)) = -\frac{\phi(\Phi^{-1}(P(Z)))}{P(Z)}$ and $\hat{\lambda}_0(P(Z)) = \frac{\phi(\Phi^{-1}(P(Z)))}{1-P(Z)}$ for the Propensity score, $P(Z)$. If γ_{12} and γ_{02} differed significantly from zero, this would provide evidence that the pattern of selection on unobservables, which in the parametric model is governed by $\hat{\lambda}_1(P(Z))$ and $\hat{\lambda}_0(P(Z))$, varies by X. We estimate the parameters in (13) and (14) jointly in a single equation that interacts the regressors in (13) with D and the regressors in (14) with $1 - D$. The results of these tests, presented in Table 10, provide reassurance that an additively separable MTE function does not seem overly restrictive in our setting.

TABLE 10—TESTS OF ADDITIVE SEPARABILITY

	Math	Spanish
F Statistic	5.59	2.23
p - value	0.908	0.984

Note: The table displays the results of tests of additive separability. P-values are obtained by 250 bootstrap replications.

V. Counterfactuals

The pattern of negative sorting on gains together with an MTE function that is positive nearly everywhere suggests that policies that expand access to telesecundarias will raise test scores in Mexico in relation to the status quo. We therefore evaluate the effects of two counterfactual policies that are likely to induce more children with high latent resistance to treatment (high U_D) to switch from traditional secondary schools to telesecundarias.

The effects of counterfactual policies can be evaluated by integrating the MTE with respect to the probability distribution induced by the proposed policy. A baseline policy, a, is characterized by a particular distribution of the instrument, Z^a. The move from policy a to a new policy a' corresponds to a shift in the distribution of the instrument from F_{Z^a} to $F_{Z^{a'}}$. This shift induces some students to attend telesecundarias who would not otherwise attend. The treatment effect for students induced to switch attendance from traditional schools to telesecundarias as a result of the policy is given by $PRTE_{a',a}$, which is positive if these students learn more at telesecundarias than at traditional schools.

A. Increasing Access to Telesecundarias

We consider a class of counterfactual policies that expand access to telesecundarias so that the relative distance to them decreases for all students, $Z^{a'} \leq Z^{a}$. As our sample omits individuals who drop out between the sixth and seventh grades, we will not be able to say anything about the distribution of test scores for students who are induced to attend telesecundarias instead of dropping out as a result of the counterfactual policy. Our PRTE estimates apply only to the population of students who were already attending secondary school under the baseline policy in 2008. We measure only the achievement effects, not the enrollment effects, of each policy.[18] We acknowledge that enrollment effects may be significant, although they are bounded above by the 5.7% of students who dropped out between the sixth and seventh grades in 2008 (see Table B-2 in Appendix B).

We consider two counterfactual policies. The first is a hypothetical policy that reduces the relative distance to telesecundarias by five km for every student. This policy, which would entail moving traditional schools farther away for students whose nearest telesecundaria is under five km, is infeasible but provides an instructive example of the gains to a policy that can drastically raise telesecundaria attendance.

The second counterfactual is a feasible school-building policy that constructs a telesecundaria directly adjacent to the twenty-two percent of primary schools that have no telesecundaria within a five km radius. This has the effect of reducing the distance between primary schools and telesecundarias to zero for all students who formerly had only a distant telesecundaria.

Table 11 presents estimates of the PRTEs for both math and Spanish alongside the IV estimate from a Two-Stage Least Squares regression with relative distance to telesecundarias as the excluded instrument.[19] All parameters are precisely estimated and statistically significant at conventional levels of significance. We find that the first, hypothetical, policy causes a 37.3 point increase in mean math scores and a 25.0 point increase in mean Spanish scores in the seventh grade. The feasible school-building policy also causes improvements, but they are smaller, 23.6 points for math and 17.7 points for Spanish, owing to the less dramatic nature of the policy. The IV estimates for both math and Spanish lie in the middle of the treatment effects of the two counterfactual policies.

The results underscore the positive academic gains that can result from policies that make telesecundarias more accessible. In addition, the considerable variation across our many estimated treatment effects underscore how, in a setting with

[18]A revealed preference argument demonstrates that no students will transition from dropping out to attendance in traditional schools as a result of the policy. The counterfactual policy, however, may induce students to transition from dropout to telesecundarias. The treatment effects we estimate therefore pertain to the set of students attending secondary school under the status quo.

[19]Figure B-2 in Appendix B displays the probability distributions of U_D corresponding to the two PRTE estimates and IV.

TABLE 11—POLICY-RELEVANT TREATMENT EFFECTS

	Hypothetical	Feasible	IV
Math	37.3	23.6	30.4
	(2.13)	(3.43)	(1.48)
Spanish	25.0	17.7	18.7
	(2.18)	(3.70)	(1.47)

Note: The table displays treatment parameters corresponding to two counterfactual policies discussed in section V The hypothetical policy reduces relative distance between telesecundarias and traditional secondary schools by five km, while the feasible policy constructs telesecundarias adjacent to all primary schools that do not have a telesecundaria within a five km radius. Both policy-relevant treatment effects are calculated using the semiparametric MTE combined with the simulation method of Carneiro, Lokshin and Umapathi (2017). Standard errors in parentheses are obtained by 250 bootstrap replications. The IV estimates are obtained by a Two-Stage Least Squares regression that uses relative distance as an instrument for telesecundaria attendance.

heterogeneous treatment effects, each policy generates a unique causal effect based on who the policy induces into treatment.

B. *Effects of Telesecundarias on Educational Inequality*

Many societies are characterized by deep cleavages, of geography, race, or language. In what follows we explain how we use the MTE to evaluate the effect of telesecundarias on two important divisions in Mexico: the test score gaps between rural and urban students and between those who speak Spanish at home and those who speak an indigenous language at home.

Let the random variable $G \in \{0, 1\}$ denote the subdivision to which an individual belongs (rural versus urban, for example). We estimate the effect of telesecundarias on the test score gap for the division defined by G by considering a counterfactual policy that eliminates telesecundarias. The causal effect of telesecundarias on the test score gap, TE^G, is the difference between the gap under the status quo policy, a, and the new policy, a':

(15)
$$TE^G = \mathbb{E}[Y^a \mid G = 1] - \mathbb{E}[Y^a \mid G = 0] - \{\mathbb{E}[Y^{a'} \mid G = 1] - \mathbb{E}[Y^{a'} \mid G = 0]\} \ .$$

We can write TE^G as a function of the average treatment effect on the treated and the probability of obtaining treatment for each separate subgroup as follows:

$$
\begin{aligned}
TE^G =& \mathbb{E}[Y_0 + D^a(Y_1 - Y_0) \mid G = 1] - \mathbb{E}[Y_0 + D^a(Y_1 - Y_0) \mid G = 0] - \\
& \{\mathbb{E}[Y_0 + D^{a'}(Y_1 - Y_0) \mid G = 1] - \mathbb{E}[Y_0 + D^{a'}(Y_1 - Y_0) \mid G = 0]\} \\
=& \mathbb{E}[D^a(Y_1 - Y_0) \mid G = 1] - \mathbb{E}[D^a(Y_1 - Y_0) \mid G = 0] - \\
& \{\mathbb{E}[D^{a'}(Y_1 - Y_0) \mid G = 1] - \mathbb{E}[D^{a'}(Y_1 - Y_0) \mid G = 0]\} \\
=& \mathbb{E}[D^a(Y_1 - Y_0) \mid G = 1] - \mathbb{E}[D^a(Y_1 - Y_0) \mid G = 0] \\
=& \mathbb{E}[Y_1 - Y_0 \mid G = 1, D^a = 1]\mathbb{P}(D^a = 1 \mid G = 1) - \\
& \mathbb{E}[Y_1 - Y_0 \mid G = 0, D^a = 1]\mathbb{P}(D^a = 1 \mid G = 0)\ , \qquad (16)
\end{aligned}
$$

where the third equality follows because $D^{a'} = 0$ for everyone when telesecundarias are eliminated, and the fourth equality follows from the law of iterated expectations. Equation (16) makes clear that telesecundarias can reduce the test score gap in two ways: through unequal treatment effects across different values of G or through differential treatment probabilities.

Table 12 presents estimates of TE^G on the gaps between urban and rural students and native Spanish speakers and those who speak an indigenous language at home. The table shows that the urban-rural test score gap in math is 10.8 points, or 0.107 standard deviations, but that it would be 13.8 points larger (128%) if telesecundarias were eliminated. The corresponding gap for seventh grade Spanish scores is 18.7 points, or 0.187 standard deviations, and it would be 8.1 points larger (43%) in the absence of telesecundarias. These differences are statistically significant. Telesecundarias instead cause smaller reductions in the test score gaps between Spanish- and Indigenous-language students, but the effect on the math score gap, at 3.1 points, is still significant.

TABLE 12—EFFECTS ON TEST SCORE GAPS

	Math		Spanish	
	Gap	Treatment Effect	Gap	Treatment Effect
Urban - Rural	10.8	13.8	18.7	8.10
		(1.03)		(1.01)
Spanish - Indigenous	15.5	3.10	22.7	1.20
		(0.99)		(0.92)

Note: The table shows the effects of telesecundarias on reducing the test score gaps between urban and rural residents and between students who speak Spanish at home and those who speak an indigenous language at home. The dependent variables are raw scores on the seventh grade math and Spanish ENLACE exams. A positive treatment effect indicates that the test score gap would be larger in the absence of telesecundarias. The treatment effects are obtained by integrating the semiparametric MTE with respect to $f(X, U_D \mid G)$ for G equal to the urban, rural, Spanish-speaking, and indigenous-speaking subsamples, respectively. The simulation method of Carneiro, Lokshin and Umapathi (2017) is used to integrate the MTE. Standard errors are obtained by 250 bootstrap replications.

VI. Conclusion

In this paper we evaluate the effectiveness of a long-running distance learning program in Mexico. Our empirical approach, which combines value-added modeling with the MTE framework, simultaneously allows for rich heterogeneity in treatment effects and lets us isolate the degree of student learning that is directly attributable to attendance in telesecundarias. We find evidence of considerable heterogeneity in value added in math and Spanish caused by attending telesecundarias but that nearly all students benefit. The gains are large and correspond to a 0.348 standard deviation increase in math scores and a 0.225 standard deviation increase in Spanish scores after a single year of attendance. Our counterfactual simulations suggest that telesecundaria expansions would yield positive academic dividends, with the magnitude of these policy effects varying with the degree to which the policies make telesecundarias more accessible. Furthermore, we show that, by providing a high-quality alternative to rural students with initially low academic skills, telesecundarias reduce the rural-urban test score gap. Since telesecundarias are cheaper than traditional schools, the benefits of sending a child to a telesecundaria instead of a traditional school would exceed the costs regardless of the assumptions used in any cost-benefit analysis.

While COVID-19 caused many school districts to hastily move existing lesson plans online and students to log into class from home, telesecundarias show that distance education can take different forms than the pandemic model. Based on the findings presented in this paper, we believe that the telesecundaria model can be an effective learning technology in areas where local educational alternatives are poor. A combination of high quality video instruction, in-person supervision, and standardized teaching materials may benefit students in these environments. Such a technology, which could be offered at a low per-unit cost by either governments or private entities, likely causes additional benefits not analyzed in this paper. We've provided evidence that, by raising knowledge, telesecundarias also raise educational attainment. Standard theories of human capital, as in Becker (2009) and Ben-Porath (1967), predict that this would have causal effects on earnings later in life. Finally, by building skills in isolated regions of Mexico, telesecundarias may help to raise a new generation of teachers that can alleviate the shortage this program was initially designed to address.

REFERENCES

Abdulkadiroğlu, Atila, Parag A Pathak, Jonathan Schellenberg, and Christopher R Walters. 2020. "Do Parents Value School Effectiveness?" *American Economic Review*, 110(5): 1502–39.

Acta Educativa, Universidad Abierta. 2016. "La Telesecundaria vs Secundarias Generales y Técnicas."

https://revista.universidadabierta.edu.mx/2016/11/30/la-telesecundaria-vs-secundarias-generales-y-tecnicas/, Accessed 2021-09-24.

Ainsworth, Robert, Rajeev Dehejia, Cristian Pop-Eleches, and Miguel Urquiola. 2022. "Why Do Households Leave School Value Added "On the Table"? The Roles of Information and Preferences." Working Paper.

Almagor, Lelac. 2021. "I Taught Online School This Year. It Was a Disgrace." *https://www.nytimes.com/2021/06/16/opinion/remote-learning-failure.html*, Accessed: 2012-01-14.

Altonji, Joseph G, Todd E Elder, and Christopher R Taber. 2005. "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling." *Journal of Human Resources*, 40(4): 791–821.

Avitabile, Ciro, and Rafael De Hoyos. 2018. "The Heterogeneous Effect of Information on Student Performance: Evidence from a Randomized Control Trial in Mexico." *Journal of Development Economics*, 135: 318–348.

Becker, Gary S. 2009. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* University of Chicago press.

Beg, Sabrin, Waqas Halim, Adrienne M Lucas, and Umar Saif. 2022. "Engaging Teachers with Technology Increased Achievement, Bypassing Teachers Did Not." *American Economic Journal: Economic Policy*, 14(2): 61–90.

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 E Todd, and Weilong Zhang. 2021. "Prospering through Prospera: CCT Impacts on Educational Attainment and Achievement in Mexico." Penn Institute for Economic Research.

Ben-Porath, Yoram. 1967. "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy*, 75(4, Part 1): 352–365.

Bianchi, Nicola, Yi Lu, and Hong Song. 2022. "The Effect of Computer-Assisted Learning on Students' Long-Term Development." *Journal of Development Economics*, 158.

Björklund, Anders, and Robert Moffitt. 1987. "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models." *The Review of Economics and Statistics*, 42–49.

Boardman, Anthony E, and Richard J Murnane. 1979. "Using Panel Data to Improve Estimates of the Determinants of Educational Achievement." *Sociology of Education*, 113–121.

Brinch, Christian N, Magne Mogstad, and Matthew Wiswall. 2017. "Beyond LATE with a Discrete Instrument." *Journal of Political Economy*, 125(4): 985–1039.

Bulman, George, and Robert W Fairlie. 2016. "Technology and Education: Computers, Software, and the Internet." In *Handbook of the Economics of Education*. Vol. 5, 239–280. Elsevier.

Calderoni, Jose. 1998. *Telesecundaria: Using TV to Bring Education to Rural Mexico.* Education and Technology Team, Human Development Network, World Bank.

Cameron, Stephen V, and Christopher Taber. 2004. "Estimation of Educational Borrowing Constraints using Returns to Schooling." *Journal of Political Economy*, 112(1): 132–182.

Card, David. 1995. "Using Geographical Variation in College Proximity to Estimate the Return to Schooling." *Aspects of Labor Market Behavior: Essays in Honor of John Vanderkamp*, 201–222.

Carneiro, Pedro, and James J Heckman. 2002. "The Evidence on Credit Constraints in Post-Secondary Schooling." *The Economic Journal*, 112(482): 705–734.

Carneiro, Pedro, James J Heckman, and Edward J Vytlacil. 2011. "Estimating Marginal Returns to Education." *American Economic Review*, 101(6): 2754–2781.

Carneiro, Pedro, Michael Lokshin, and Nithin Umapathi. 2017. "Average and Marginal Returns to Upper Secondary Schooling in Indonesia." *Journal of Applied Econometrics*, 32(1): 16–36.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg. 2018. "Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance." *Journal of Political Economy*, 126(6): 2356–2409.

Currie, Janet, and Enrico Moretti. 2003. "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings." *The Quarterly Journal of Economics*, 118(4): 1495–1532.

de Cossio, Roger Diaz. 2007. "Repensar la Telesecundaria." *El Universal*.

De Groote, Olivier, and Koen Declercq. 2021. "Tracking and Specialization of High Schools: Heterogeneous Effects of School Choice." *Journal of Applied Econometrics*, 36(7): 898–916.

de Hoyos, Rafael, Ricardo Estrada, and María José Vargas. 2021. "What Do Test Scores Really Capture? Evidence from a Large-Scale Student Assessment in Mexico." *World Development*, 146.

De Hoyos, Rafael, Vicente A Garcia-Moreno, and Harry Anthony Patrinos. 2017. "The Impact of an Accountability Intervention with Diagnostic Feedback: Evidence from Mexico." *Economics of Education Review*, 58: 123–140.

Duflo, Esther, Rema Hanna, and Stephen P Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review*, 102(4): 1241–78.

Dustan, Andrew, Alain De Janvry, and Elisabeth Sadoulet. 2017. "Flourish or Fail? The Risky Reward of Elite High School Admission in Mexico City." *Journal of Human Resources*, 52(3): 756–799.

Engzell, Per, Arun Frey, and Mark D Verhagen. 2021. "Learning Loss Due to School Closures during the COVID-19 Pandemic." *Proceedings of the National Academy of Sciences*, 118(17).

Escuelasmex. 2018. *Directorio de Instituciones Educativas en México*. https://escuelasmex.com/. Accessed July 20, 2019.

Escueta, Maya, Vincent Quan, Andre Joshua Nickow, and Philip Oreopoulos. 2017. "Education Technology: An Evidence-Based Review." *National Bureau of Economic Research*.

Estrada, Rafael Quiroz. 2003. "Telesecundaria: Los Estudiantes y los Sentidos que Atribuyen a Algunos Elementos del Modelo Pedagógico." *Revista Mexicana de Investigación Educativa*, 8(17).

Estrada, Ricardo. 2019. "Rules Versus Discretion in Public Service: Teacher Hiring in Mexico." *Journal of Labor Economics*, 37(2): 545–579.

Estrada, Ricardo, and Jérémie Gignoux. 2017. "Benefits to Elite Schools and the Expected Returns to Education: Evidence from Mexico City." *European Economic Review*, 95: 168–194.

Fabregas, Raissa. 2020. "Broadcasting Education: The Long-Term Effects of Mexico's Telesecundarias." *Working Paper*.

Fan, Jianqing, and Irene Gijbels. 1996. *Local Polynomial Modelling and its Applications: Monographs on Statistics and Applied Probability*. Vol. 66, CRC Press.

Grogger, Jeffrey, and Derek Neal. 2000. "Further Evidence on the Effects of Catholic Secondary Schooling." *Brookings-Wharton Papers on Urban Affairs*, 151–201.

Heckman, James J, and Edward J Vytlacil. 1999. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects." *Proceedings of the National Academy of Sciences*, 96(8): 4730–4734.

Heckman, James J, and Edward J Vytlacil. 2001*a*. "Local Instrumental Variables." In *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya.* , ed. K. Morimune C. Hsiao and J. L. Powell, 1–46. New York:Cambridge University Press.

Heckman, James J, and Edward J Vytlacil. 2001*b*. "Policy-Relevant Treatment Effects." *American Economic Review*, 91(2): 107–111.

Heckman, James J, and Edward J Vytlacil. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica*, 73(3): 669–738.

Heckman, James J, and Edward J Vytlacil. 2007. "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments." *Handbook of Econometrics*, 6: 4875–5143.

Johnston, Jamie, and Christopher Ksoll. 2017. "Effectiveness of Interactive Satellite-Transmitted Instruction: Experimental Evidence from Ghanaian Primary Schools. CEPA Working Paper No. 17-08." *Stanford Center for Education Policy Analysis*.

Kane, Thomas J, and Cecilia Elena Rouse. 1995. "Labor-Market Returns to Two- and Four-year College." *The American Economic Review*, 85(3): 600–614.

Kline, Patrick, and Christopher R Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *The Quarterly Journal of Economics*, 131(4): 1795–1848.

Kling, Jeffrey R. 2001. "Interpreting Instrumental Variables Estimates of the Returns to Schooling." *Journal of Business & Economic Statistics*, 19(3): 358–364.

Maldonado, Joana Elisa, and Kristof De Witte. 2021. "The Effect of School Closures on Standardised Student Test Outcomes." *British Educational Research Journal*.

Michaelsen, Maren M, and Paola Salardi. 2020. "Violence, Psychological Stress and Educational Performance During the "War on Drugs" in Mexico." *Journal of Development Economics*, 143.

Navarro-Sola, Laia. 2021. "Secondary Schools with Televised Lessons: The Labor Market Returns of the Mexican Telesecundaria." *Working Paper*.

Neal, Derek. 1997. "The Effects of Catholic Secondary Schooling on Educational Achievement." *Journal of Labor Economics*, 15(1, Part 1): 98–123.

Parente, Paulo MDC, and JMC Santos Silva. 2012. "A Cautionary Note on Tests of Overidentifying Restrictions." *Economics Letters*, 115(2): 314–317.

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

Robinson, Peter M. 1988. "Root-N-Consistent Semiparametric Regression." *Econometrica*, 931–954.

Rodriguez-Segura, Daniel. 2020. "Educational Technology in Developing Countries: A Systematic Review." *Working Paper*.

Roy, Andrew Donald. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers*, 3(2): 135–146.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology*, 66(5): 688.

Santos, Annette. 2001. "Oportunidades Educativas en Telesecundaria y Factores que las Condicionan." *Revista Latinoamericana de Estudios Educativos (México)*, 31(3): 11–52.

Sapelli, Claudio, and Bernardita Vial. 2002. "The Performance of Private and Public Schools in the Chilean Voucher System." *Cuadernos de Economía*, 39(118): 423–454.

SEP. 2018*a*. "Cuestionarios de Contexto (ENLACE)." *Secretaría de Educación Pública*. www.xaber.org.mx. Accessed November 15, 2018.

SEP. 2018*b*. "Resultados de Aprendizaje (ENLACE)." *Secretaría de Educación Pública*. www.xaber.org.mx. Accessed November 15, 2018.

Todd, Petra E, and Kenneth I Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *The Economic Journal*, F3–F33.

Walters, Christopher R. 2018. "The Demand for Effective Charter Schools." *Journal of Political Economy*, 126(6): 2179–2223.