Quantitative Economics 16 (2025), 133–183 1759-7331/20250133

Prospering through *Prospera*: A dynamic model of CCT impacts on educational attainment and achievement in Mexico

Jere R. Behrman
Department of Economics, University of Pennsylvania

Susan W. Parker
School of Public Policy, University of Maryland

Petra Todd
Department of Economics, University of Pennsylvania, NBER, HCEO, and IZA

Weilong Zhang
Faculty of Economics, University of Cambridge

This paper develops and estimates a dynamic model, which integrates value-added and school-choice models, to evaluate grade-by-grade and cumulative impacts of the Mexican *Prospera* conditional cash transfer (CCT) program on educational achievement. The empirical application advances the previous literature by estimating policy impacts on learning, accounting for dynamic selective school attendance, and incorporating both observed and unobserved heterogeneity. A dynamic framework is critical for estimating cumulative learning effects because lagged achievements are important determinants of current achievements. The model is estimated using rich nationwide Mexican administrative data on schooling progression and mathematics and Spanish test scores in grades 4–9 along with student and family survey data. The estimates show significant CCT impacts on learning and educational attainment, particularly for students from poorer households. Results show that telesecondary schools (distance learning) play a crucial role in facilitating school attendance and in fostering skill accumulation.

Keywords. Conditional cash transfers, dynamic modeling, educational attainment, learning achievement, Mexico.

Jere R. Behrman: jbehrman@econ.upenn.edu
Susan W. Parker: swparker@umd.edu
Petra Todd: ptodd@econ.upenn.edu
Weilong Zhang: weilong.zhang@econ.cam.ac.uk
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." We also thank Gabrielle Vasey, Rodrigo Deiana, Elizaveta Brover, Pinar Goktas, Mira Potter-Schwartz, and Erika Trevino-Alvarado for research assistance. We thank Hector Robles Vasquez for preparing databases and for assistance in working with these data. We thank Miguel Szekely for conversations about the Mexican educational system. This paper was presented at the University of Cambridge, the University of Arizona, the University of Maryland, McGill University, the University of Pennsylvania, the Stanford Institute for Theoretical Economics, and the University of Glasgow. We thank Eric French, Magne Mogstad, and Christopher Taber for helpful suggestions.

Available at http://qeconomics.org. https://doi.org/10.3982/QE2291

JEL classification. C53, I25, I38, J24.

1. Introduction

Conditional cash transfer (CCT) programs aim to alleviate current poverty through transfers to poor families and to reduce future poverty by making these transfers conditional on investments in the human capital of children and youth. In 1998–2000, a large-scale randomized evaluation of the Mexican *PROGRESA* CCT program demonstrated substantial impacts on schooling enrollment and attainment, child work, and family income (Parker and Todd (2017)). These findings contributed to a large scaling-up in Mexico and an impressive adoption of similar programs in more than 60 countries on five continents (Fiszbein and Schady (2009)).

This paper analyzes the understudied, substantive question of whether CCTs improve learning by developing and estimating a dynamic model of academic achievement and school progression. Several studies, using various methods including the original experiment, matching and structural dynamic models, have examined the impacts of *PROGRESA/Oportunidades/Prospera* on school enrollment and, in some cases, on longer-term schooling attainment (e.g., Schultz (2004), Behrman, Sengupta, and Todd (2005), Behrman, Parker, and Todd (2005, 2009), Todd and Wolpin (2006), Attanasio, Meghir, and Santiago (2012), Parker and Vogl (2023)). This literature demonstrated positive program impacts on school enrollment and attainment. However, a longstanding concern has been whether and to what extent this increased school enrollment translates into higher academic achievement, a likely determinant of the extent to which CCTs can improve earnings potential and other longer-term outcomes. Most prior studies did not analyze academic achievement impacts, because the original evaluation data did not include achievement test scores.[1]

With newly available data, we are now able to examine the effects of the *Prospera* program (the program name during the time of our data collection) not only on school enrollment and attainment but also on academic achievement in mathematics and Spanish. Nationwide standardized longitudinal administrative test-score data (called the *ENLACE* data) as well as complete enrollment rosters were merged with administrative information on which students come from *Prospera* households and on school locations. They were also merged with survey information obtained from students and their parents. These data allow the study of how students' *Prospera* beneficiary status affects their school enrollment, school choice, grade progression, and academic achievements over time.

Although our longitudinal administrative and survey data are rich and have the advantages of national coverage and large sample sizes, the data were not collected explicitly for the purpose of evaluating the *Prospera* program. There are at least five significant

[1]In 2003, Woodcock–Johnson tests in mathematics and Spanish were applied to a single cross-section. Using these data, Behrman, Parker, and Todd (2009) found no impacts of *PROGRESA* participation on achievement, based on comparing test scores of the original treatment and control groups. However, because the original control group was enrolled in the program 1.5 years after the original treatment group, the schooling differences between them were relatively small, at about 0.2 years of additional schooling.

statistical challenges in using observational data of this kind to assess program impacts: (i) selective program participation, largely due to eligibility criteria that restrict access to high-poverty households, (ii) nonrandom school dropout, which mainly occurs after grade 6, (iii) grade retention in any grade, (iv) the availability of multiple school types and school choice, and (v) the presence of a small fraction of students suspected to have cheated on the tests. This paper develops a methodological approach for evaluating the effects of *Prospera* and illuminating the mechanisms through which program effects operate while accounting for these different features of the data and the context.

Some earlier studies consider the selection problem arising from nonrandom dropout in the context of analyzing educational outcome determinants (see, e.g., Cameron and Heckman (1998, 2001), Glewwe (2002)). As is common in administrative schooling data, including our data, the test scores are only observed for enrolled children who took the tests at school. The selection problem is dynamic as it occurs at each grade and the students at risk for dropping-out in a particular grade depend on the sample that stayed in school from previous grades. The *Prospera* CCT program induced students from high-poverty backgrounds who were at high risk for dropping-out to stay in school longer. If the CCT program induces weaker students to remain in school, then average test scores could fall as a result of more marginal students being included in the testing. This kind of selection problem arises whenever tests are administered in school, regardless of whether the data analyzed are experimental or nonexperimental.[2]

Our goal in this paper is to examine how the *Prospera* program affects schooling and academic achievement, accounting for the statistical problems noted in (i)–(v) in the penultimate paragraph. To this end, we develop and estimate a dynamic model of students' school progression that incorporates decision making in each grade (4–9) with regard to enrollment, school choice, and dropping-out as well as grade- and subject-varying models for academic achievement. Specifically, our modeling framework combines value-added academic achievement models with school-choice models and links equations across ages/grades, allowing for both observed and unobserved heterogeneity. Valued-added models typically specify a relationship among academic achievement, key learning inputs in the current period, and lagged achievement, which is a sufficient statistic for past learning inputs under some assumptions about coefficients of past learning inputs following geometric patterns (Summers and Wolfe (1977), Boardman and Murnane (1979), Hanushek (1979), Todd and Wolpin (2003), Cunha, Heckman, Lochner, and Masterov (2006), Cunha, Heckman, and Schennach (2010)).[3] School-choice models generally focus on the decision of what type of school to attend (Neal (1997), McEwan (2001), Altonji, Elder, and Taber (2005), Sapelli and Vial (2002), Gallego

[2]This selection problem also affects cross-country comparisons of standardized tests, such as PISA test scores. The PISA tests are given in schools at age 15 and, in some countries, significant fractions of children have dropped-out by that age.

[3]There is some debate about whether value-added models with teacher fixed effects should be used to measure teacher effectiveness (see, e.g., Kane and Staiger (2008), Kane, McCaffrey, Miller, and Staiger (2013), Chetty, Friedman, and Rockoff (2014a,b)). Our focus is rather on using these models to capture the cumulative learning process.

and Hernando (2009)).[4] Value-added models and school-choice models are usually estimated in isolation, although there are a few papers that combine them (e.g., Hastings, Neilson, and Zimmerman (2012), Allende et al. (2019), Schellenberg and Walters (2020)). Our model also incorporates dropout decisions and grade retention.

The model begins when students finish 4th grade and continues through the 9th grade.[5] Students in our sample differ in terms of their family backgrounds and their 4th-grade knowledge in mathematics and Spanish as measured by standardized test scores. The vast majority of children attend general primary schools close to home, but some families living in areas with high indigenous populations can choose between general and bilingual indigenous schools. At the end of each grade, students can progress to the next grade, repeat the same grade or (after grade 6), drop out. Conditional on progressing to lower-secondary school (at the end of grade 6), students/parents make a one-time choice of a lower-secondary school type, from up to three public school options—general, telesecondary, or technical schools—all of which are academically oriented.[6]

In each period, we model skill accumulation in mathematics and Spanish using value-added production functions, with coefficients that vary by grade, school types, and grade-retention status. The dynamic-panel specification captures the notion that skill accumulation at one stage affects skill attainment at other stages, which has been shown to be essential to characterizing human-capital-skill formation processes (e.g., Cunha et al. (2006), Cunha, Heckman, and Schennach (2010)). When students/parents choose from among different school types, they essentially choose a learning technology. The same inputs (including *Prospera* participation) may generate substantially different outcome trajectories depending on the school types that are available and are selected.

As previously noted, the model we estimate controls for selection arising from school-enrollment, dropout, grade-retention, and school-type choices, all of which potentially affect students' grade progression and academic achievements. It also accounts for selective program participation arising from the fact that only high-poverty households are eligible to participate in *Prospera*. In particular, we limit our analysis sub-

[4]Some studies use school-choice models to estimate school-voucher effects (Rouse (1998), Figlio and Elena Rouse (2006), Hsieh and Urquiola (2006), Bravo, Mukhopadhyay, and Todd (2010), Angrist, Bettinger, Bloom, King, and Kremer (2002), Angrist, Bettinger, and Kremer (2006)), to study parents' preferences for school quality, and to analyze the welfare effects of school policies (Epple, Jha, and Sieg (2018), Hastings, Kane, and Staiger (2009), Allende et al. (2019)).

[5]The *Prospera* cash transfers for attending school actually start in grade 3. We start our model at grade 4, because the data were collected over a 6-year time frame and we want to follow students through grade 9, which is the last grade of lower-secondary school. There are no nationwide standardized tests in the 10th and 11th grades. By starting the model at grade 4, we do not capture potential program impacts in prior grades and potentially understate program benefits.

[6]Technical schools differ from general schools by including vocational/technical educational curricular components. Telesecondary schools are distance-learning schools that largely serve rural communities and that enroll almost 20% of lower-secondary school students. *Prospera*-beneficiary family children attend telesecondary schools in greater proportions than average. We exclude private schools from the choice set as almost no Progresa beneficiaries attend private school. Section two provides more detail on how the school types differ.

sample to households that are estimated to have positive probability of being a program beneficiary and, in addition, control for observed heterogeneity between *Prospera* and non-*Prospera* students using a rich set of family demographics. Our data set contains information on most of the variables used to determine *Prospera* eligibility, but we also allow for the possibility of selection on some unobserved factors.[7] The unobserved heterogeneity is modeled as discrete latent multinomial types, as in Heckman and Singer (1984), Cameron and Heckman (1998). These types enter multiple model equations and, in doing so, allow for across-equation correlated error structures. We allow the unobserved-type distribution to vary by *Prospera* beneficiary status and an index measure of local poverty.

The data we analyze also contain information on small percentages of students in each grade that are suspected to have copied answers on the multiple-choice standardized tests. In the context of a value-added test-score model, copying induces one-sided measurement errors in the dependent variables, the lagged dependent variables, or both, which we explicitly take into account in our maximum likelihood estimation procedure. The outcomes at different ages/grades are school enrollment, school choices, mathematics, and Spanish test scores, dropping-out, and grade retention. We do not know of previous research that estimates value-added models accounting for possible cheating, although cheating on standardized tests is a ubiquitous problem.

We use our estimated model framework to evaluate how *Prospera*-beneficiary status affects schooling progression and academic achievements in different grades. In particular, we simulate school-choice decisions, school-enrollment decisions, and academic achievements with and without the *Prospera* program, for children from different family backgrounds. There are multiple channels through which *Prospera* participation can affect these outcomes. First, past participation may increase lagged achievement, which can facilitate present learning. For example, greater comprehension of 6th-grade mathematics can facilitate learning and comprehension of the 7th-grade curriculum.[8] Second, contemporaneous program participation can directly affect learning if the program encourages regular school attendance, student engagement, and study efforts. There are two reasons why we might expect the *Prospera* program to influence students in this way. *Prospera* program rules stipulated that children must attend school at least 85% of days and can only fail a grade once to receive the cash transfers. Additionally, *Prospera* transfers may reduce the pressure on children/youth to work in labor markets while in school and thereby allow for greater focus on schoolwork (Skoufias and Parker (2001)).

Our analysis yields a number of findings regarding *Prospera*-program effects and the effectiveness of different school types. First, we find that *Prospera* participation reduces lower-secondary school dropout rates by 0.06–0.09 percentage points. This effect is most

[7]As discussed in Todd and Wolpin (2003), Rivkin, Hanushek, and Kain (2005) for value-added models and in numerous other studies of schooling (e.g., Behrman, Hrubec, Taubman, and Wales (1980), Behrman and Rosenzweig (1999), Altonji, Elder, and Taber (2005), Rothstein (2009)), it is important to control for unobserved inherent student abilities, personality traits, or motivation, that matter for children's achievement growth.

[8]Cunha et al. (2006) term this feature of cognitive achievement production functions "self-productivity."

pronounced during the transition from 6th to 7th grade and appears similar for both girls and boys.[9] In primary-school grades (grades 5 and 6), our findings do not show any significant impact of the program on test scores. However, in lower-secondary grades (grades 7 to 9), we observe positive and statistically significant improvements in test scores. The effects are more substantial in mathematics, with a cumulative effect of 0.21 standard deviations by the 9th grade, compared to 0.04 in Spanish. Girls tend to have higher gains in mathematics, whereas boys show greater gains in Spanish. Therefore, the *Prospera* program narrows existing gender test score gaps that favor boys in mathematics and girls in Spanish. Additionally, and as expected, the effects of program participation accumulate over time and intensify with prolonged exposure. Intriguingly, children from more disadvantaged backgrounds exhibit significantly larger improvements in test scores.

Our results contribute to the understudied yet substantive question of whether CCTs improve learning. Fiszbein and Schady (2009), Baird, Ferreira, Özler, and Woolcock (2014) systematically reviewed a range of CCT programs worldwide and concluded that the effects of CCTs on achievement tests were disappointingly "small, at best." One caveat is that these conclusions are mostly drawn from relatively short evaluation periods and based on relatively small sample sizes.[10] When evaluating CCT programs over longer horizons, some studies report statistically significant effects on academic achievement. For example, Barham, Macours, and Maluccio (2013) used the randomized phase-in of the *Red de Proteccion Social* CCT program in Nicaragua to study effects on schooling attainment and learning for boys 10 years later. They found a half-grade increase in schooling and substantial gains (approximately 0.25 standard deviations) in mathematics and language achievement scores. Comparing two cohorts (2007 and 2013), Hadna and Kartika (2017) found statistically significant effects of a CCT program called *Program Keluarga Harapanin* in Indonesia on three subjects (Bahasa Indonesia, mathematics, and English) as well as national mathematics examinations for junior/high-school students. Our results can reconcile some of the mixed findings in the literature by highlighting the accumulative feature of the CCT program effects.

Our study sheds light on the efficacy of various types of Mexican schools in enhancing test scores. This includes telesecondary schools, which predominantly utilize video-based teaching methods and are often the only accessible option for students in rural areas. Our findings indicate that telesecondary schools are, in many instances, as effective or even more so than regular public schools for their attendees. When we use the estimated model to analyze the effect of removing the telesecondary option from the choice set, we find that the dropout rate would be substantially higher and average schooling attainment lower without these schools. Model simulations also show that telesecondary schools are also important determinants of *Prospera* program impacts. Our finding that these schools play an important role in fostering education in

[9]Some studies have suggested a greater impact of *PROGRESA* on secondary-school enrollment for girls (Schultz (2004), Parker and Vogl (2023)), but other research finds similar effects for both genders in lower-secondary education (Behrman, Sengupta, and Todd (2005), Todd and Wolpin (2006)).

[10]Due to data limitations, there are many fewer studies of achievement than there are of enrollment. For instance, Snilstveit et al. (2017) reviewed 38 studies of the effects of transfers on enrollment; only 11 of the programs analyzed effects on achievement.

Mexico is consistent with the difference-in-difference analysis of Fabregas and Navarro-Sola (2023) that found that the expansion of telesecondary schools led to substantial increases in schooling attainments for local students.[11]

Lastly, we also explore how inferences based on our model depend on specification assumptions. As a benchmark, we estimate a simpler value-added model grade-by-grade, without controlling for selection from multiple sources (dropout, school choice, grade retention). Comparing the results to those derived from our richer model, the cumulative program impacts are noticeably smaller. Thus, failing to control for dynamic selection would lead to underestimation of *Prospera's* impact. We identify three potential reasons for downward biases. First, the program causes students at the margin of dropping-out to stay in school longer and failure to control for this changing composition of students would lead to a downward bias in the impact estimates. Second, the simpler model does not allow heterogeneous impacts across different types of schools and, therefore, does not capture that telesecondary schools are particularly effective for *Prospera* beneficiaries. Lastly, the simpler model ignores the negative selection of unobserved types, which also leads to an underestimation of program impacts. We find that a richer modeling framework is required to capture heterogeneous program impacts and to control for multiple sources of selection bias.

The paper develops as follows. Section 2 briefly describes the Mexican school system and the data sets used in this study. Section 3 describes the model and Section 4 the estimation approach. Section 5 presents the empirical results. Section 6 presents the estimated cumulative *Prospera* program effects. It also performs model simulations where the telesecondary schooling option is removed, examines the robustness of program impact estimates to alternative modeling assumptions and explores longer-term program impacts (up to grade 12). Section 7 concludes. The Supplemental Appendix (Behrman, Parker, Todd, and Zhang (2024)) provides additional details on data sources and complete model estimates. The Supplemental Appendix will be referred to throughout the paper as SA.

2. Background

2.1 *Mexican educational system and child-labor laws*

The Mexican educational system consists of three levels: primary, secondary, and tertiary education. Formal basic education includes preschool, primary school (grades 1–6), and lower-secondary school (grades 7–9), all of which are compulsory. However, compulsory schooling laws are not well enforced. Many children dropout before completing grade 9, particularly children from lower-SES families, indigenous backgrounds, and rural areas.

Our analysis focuses on public schools. Although *Prospera* beneficiaries may choose which school to attend, in practice, almost all attend public schools (in our data only

[11]A recent paper by Borghesan and Vasey (2024) also studies the effectiveness of Mexican telesecondary schools using a marginal treatment effects (MTE) estimation approach applied to a Roy model and using the same data we analyze but focusing on 7th graders.

0.28% of beneficiaries in 6th grade are enrolled in private schools). Public primary and lower secondary schools, as part of "educacion basica," are free of charge. The Secretariat of Public Education (SEP) standardizes curriculum content, which includes Spanish, mathematics, natural sciences, history, geography, art, and physical education.[12] Secondary school is divided into lower-secondary school (grades 7–9) and upper-secondary school (grades 10–12). Lower-secondary school is free and students may follow either a general academic track or a technical track, which has more of a vocational focus. Both tracks are designed to prepare students for further education. There are fewer lower-secondary schools than primary schools and attending lower-secondary schools often requires traveling some distance from home, particularly for children living in more remote areas. Public schools do not generally provide transportation. Upper-secondary education (grades 10–12) did not become compulsory until 2012. Some upper-secondary schools are affiliated with large public universities, while others are SEP or state-controlled. At the tertiary level, the Mexican educational system has many different programs and degree options.

The Mexican Constitution prohibits child labor for minors under 14 years of age. However, the child-labor laws are not well enforced. 8% of children age 12 report working for pay in the 2010 Mexican census data.[13]

2.2 *ENLACE test-score data and additional survey data*

From 2006 to 2013, the SEP applied the *Evaluación Nacional de Logro Académico en Centros Escolares*, called the *ENLACE* (SEP (2018a)). The test evaluated student performance in mathematics, Spanish, and a rotating subject for all 3rd-to-9th graders at the end of each academic year. The test is directly based on the curriculum (see SEP (2010)) and intended to be an assessment that is informative about learning outcomes to SEP and to parents. In primary school and in lower-secondary school, the grades studied in this paper, the test has no bearing on students' GPA or grade progression, so it can be considered to be low stakes. Beginning in 2008, *ENLACE* was also given to students in their final year of upper-secondary school (grade 12).[14] The exams were designed to have a mean of 500 and a standard deviation of 100 in their first year of implementation, and subsequent test years were calibrated to allow measurement of changes in learning over time (see SEP (2010)). The test-completion rate is close to 90%. As described by De Hoyos, Estrada, and Vargas (2018), 15.1 million students in 136,000 schools took the examination in 2013, the last year the test was applied. In addition to test scores, the ENLACE data (merged with school roster data) also contain information on the age, gender, *Prospera* beneficiary status, whether the child attended the day of the test, school ID, and school type for each student. We examine a cohort of students who were in grade 4 in

[12] The National Institute for Assessment of Education (INEE) monitored standards during our period of study.

[13] Based on the authors' tabulations.

[14] The ENLACE exams have been used as a means of evaluating educational interventions by several papers (Avitabile and De Hoyos (2018), De Hoyos Navarro, Attanasio, and Meghir (2019), De Hoyos, Garcia-Moreno, and Patrinos (2017)). Scores on these exams have been shown to have predictive power on important life outcomes including university enrollment and wages (De Hoyos, Estrada, and Vargas (2018)).

2007/2008 for whom ENLACE test scores are available for up to 6 years (i.e., from 4th grades through 9th grades for students who progressed one grade each year).

We link the ENLACE data with survey information on student, parent, and school characteristics that was obtained from a random sample of schools each year (SEP (2018b)). The survey data provide detailed information on parental schooling, monthly family income, home infrastructure, number of siblings, and other household characteristics. The combination of the ENLACE data with the student/parent/school surveys produces an exceptionally rich data set that is representative of Mexican school-age children. The detailed survey information available on housing characteristics is useful for reliably approximating the *Prospera* program eligibility criteria, as will be discussed below.

In addition, we gather geocode information (latitude and longitude) for each primary and secondary school. We use such information to characterize the number of local primary schools (within 5 km radius) and number of local secondary schools (within 10 km radius). We then further calculate the distance (in kilometers) between the primary school and the nearest secondary school of each type. Some further details are provided in SA Section A.

2.3 *Descriptive statistics*

Table 1 shows summary statistics for the students in our sample. The columns show the means and standard deviations for children whose families are *Prospera* beneficiaries or nonbeneficiaries. The average age of the children is around 10 and 49% are female. The parents of beneficiary students have much less schooling; 65% of the fathers and 68% of the mothers in beneficiary families have primary school (6 grades) or less in comparison to 27% and 31% for nonbeneficiaries. Beneficiary fathers are less likely to work full-time, whereas mothers are less likely to work in the labor market at all. Beneficiary students are slightly more likely to have mothers who do not live at home. The two groups differ substantially in terms of geographic residence; 89% of nonbeneficiary students live in urban areas in comparison to 44% for beneficiaries.

Prospera families tend to be larger, with 39% having six or more household members in comparison to 24% for nonbeneficiaries. Children from beneficiary families have fewer years of preschool education on average. They are also much less likely to have access to computers or the internet at home. 17% of beneficiary students have computers at home and 11% have access to the internet in comparison to 45% and 31% for nonbeneficiaries. In terms of languages spoken at home, 4% of children from beneficiary families speak indigenous languages (sometimes in combination with Spanish) in comparison to 1% of nonbeneficiary children. Also, a third of the *Prospera*-beneficiary children live in southern states that are relatively poor compared to less than one-fourth for nonbeneficiaries.

Figure 1 shows the average test scores by grade and by *Prospera*-beneficiary status, where the solid line denotes that the family participates in *Prospera*. In 4th grade, there are considerable gaps in average mathematics and Spanish test scores. Comparing the test-score distributions across grades, we see a pattern of diminishing test-score disparities between *Prospera* and *non-Prospera* students. By grades 8 and 9, *Prospera* and

Table 1. Descriptive statistics.

	Prospera		Non-*Prospera*			*Prospera*		Non-*Prospera*	
	Mean	Std	Mean	Std		Mean	Std	Mean	Std
Age (at grade 4)	10.06	0.69	9.88	0.56	Number of household members				
Female	0.49	0.50	0.50	0.50	*≤3 people*	0.21	0.41	0.33	0.47
Schooling cat. (dad)					*4 people*	0.21	0.41	0.27	0.44
Below primary school	0.41	0.49	0.13	0.33	5 people	0.19	0.39	0.15	0.36
Primary school completed	0.24	0.43	0.14	0.34	≥6 people	0.39	0.49	0.24	0.43
Secondary or below	0.26	0.44	0.37	0.48	Number of preschool years				
College or above	0.07	0.25	0.34	0.47	0 year	0.03	0.16	0.02	0.13
Schooling cat. (mom)					*1 year*	0.11	0.31	0.04	0.20
Primary school	0.42	0.49	0.14	0.35	2 years	0.26	0.44	0.24	0.43
Primary school completed	0.26	0.44	0.17	0.38	3 years	0.31	0.46	0.45	0.50
Secondary or below	0.27	0.44	0.36	0.48	4 years	0.30	0.46	0.25	0.43
College or above	0.05	0.21	0.32	0.47	Internet at home	0.11	0.32	0.31	0.46
Working status (dad)					Computer at home	0.17	0.38	0.45	0.50
Part-time	0.22	0.41	0.16	0.37	First language				
full-time	0.74	0.44	0.80	0.40	*Spanish*	0.90	0.30	0.98	0.14
Working status (mom)					*Indigenous*	0.06	0.24	0.01	0.09
Housework	0.77	0.42	0.55	0.50	*Both*	0.04	0.19	0.01	0.10
Part-time	0.13	0.33	0.28	0.45	Region				
Full-time	0.09	0.29	0.17	0.37	*North*	0.19	0.40	0.16	0.37
Father at home	0.85	0.36	0.85	0.35	*North-center*	0.33	0.47	0.45	0.50
Mother at home	0.96	0.19	0.98	0.14	*Center*	0.14	0.35	0.18	0.38
Urban dummy	0.44	0.50	0.89	0.31	*South*	0.33	0.47	0.21	0.41
Marginality dummy	0.19	0.39	0.75	0.43					
Obs	50,857		126,246		Obs.	50,857		126,246	

Note: Authors' calculations using ENLACE merged with student- and parent-context questionnaires.

non-Prospera students have similar mathematics test-score distributions but there is still a gap for Spanish test scores. The observed narrowing of the gaps may not necessarily reflect relative improvements for *beneficiaries*, however, because it may be due to selection that occurs at various stages, which we next consider.

2.4 *Selection problems at different stages of schooling*

As described in Section 1, assessing the educational impacts of the *Prospera* program requires controlling for multiple sources of selection, including selective program participation, grade retention, drop-out, school choices, and cheating. We next present evidence on the empirical relevance of each of these factors.

Prospera-program selection Households' participation in the *Prospera* program is subject to stringent eligibility criteria that restrict access to high-poverty households. The vast majority of families who are eligible to participate opt to do so, reflecting (i) the transfers that families receive under the program are large, accounting for a 20% increase in family income on average; (ii) the program provides transfers in grades 3–6 when school attendance is nearly universal; and (iii) the program has been in operation

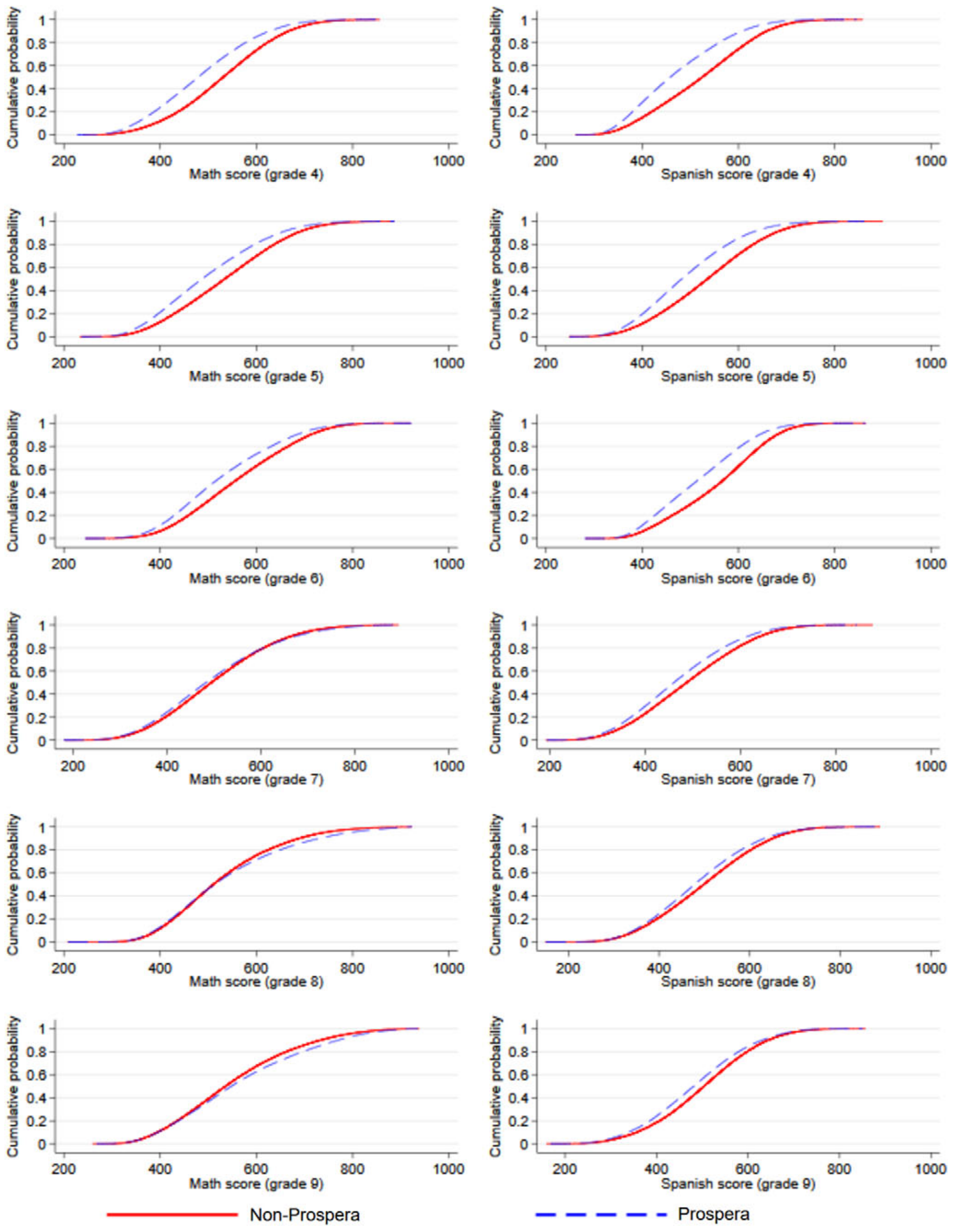

FIGURE 1. Mathematics and Spanish test-score distributions (CDFs) by grade and *Prospera* status.

since 1997 and is well known (see Parker and Todd (2017)). However, there may still be some nonparticipating families, particularly ones that are unaware of their eligibility. Also, families may receive partial benefits if they send some but not all of their children to school. We consider a student to be in a beneficiary family if they are listed in the

TABLE 2. Percentages of retained students by school type, grade, and *Prospera* status (P).

Primary school	General		Indigenous			
	$P=0$	$P=1$	$P=0$	$P=1$		
4th year	1.2%	2.5%	2.0%	3.2%		
5th year	0.8%	1.5%	1.2%	2.5%		
6th year	0.1%	0.2%	0.0%	0.1%		
	General		Telesecondary		Technical	
Lower-secondary school	$P=0$	$P=1$	$P=0$	$P=1$	$P=0$	$P=1$
1st year	0.6%	0.3%	0.9%	0.4%	0.4%	0.4%
2nd year	0.7%	0.6%	0.3%	0.3%	0.7%	0.5%

Note: This table displays the percentages of students retained by grade, school types, and Prospera-beneficiary status ($P=1$ denotes *Prospera* participation).

Prospera administrative data as being enrolled in 6th grade.[15] As was seen in Table 1, the average *Prospera* student is not directly comparable to the average *non-Prospera* student. Beneficiary students come from higher poverty backgrounds and are more likely to live in rural areas.

Grade retention Grade retention has the potential to alter the student composition within each grade cohort, another source of selection. Table 2 shows the percentage of students retained by grade, school types, and *Prospera*-beneficiary status, where $P=1$ denotes being a beneficiary, else $P=0$. It is clear that grade retention is more prevalent in primary-school grades, with notably higher rates for *Prospera* children. For lower-secondary school, a retention rate of approximately 1% is observed, with no discernible pattern by beneficiary status.

Nonrandom school dropout The Mexican Basic Education system mandates compulsory education from grades 1 to 9 and Mexican law explicitly prohibits child labor for children under the age of 14. Table 3 shows the grade distribution and school enrollment status by age. As can be seen, conditional on age, children attend a variety of grades. This grade dispersion occurs both because of delayed school enrollments and because of grade retentions. There is a pattern of increasing dropout rates with age as one might expect if the opportunity costs of attending school is to work and older children receive higher wage offers. According to the data, 5% of 13-year-olds and 10% of 14-year-olds are not currently enrolled in school. The dropout rate climbs to 17% for children aged 15. For students who remain enrolled in lower-secondary school (grades 7 to 9) at the ages of 16 and 17, half of them opt to discontinue their education. Given significant observed dropout rates and considering the fact that *Prospera* provided monetary incentives to stay in school and not drop out, it is essential that our model accounts for dropout decisions.

[15]The administrative data on participation in *Prospera*-was linked with the test-score database when children were in grade 6. The overwhelming majority of children enroll in grade 6, although there is a small fraction that drops out prior to entering grade 6 that we do not include in our analysis sample.

School type and choice In Mexico, families can choose among a few different types of primary and secondary schools, which introduces another source of selection. As previously noted, school choice is more limited in rural areas, and, for lower-secondary schooling, some families may only have access to telesecondary schools. Table 4 shows the fractions of *Prospera* beneficiary ($P = 1$) and non-*Prospera* beneficiary ($P = 0$) students attending different types of schools. Only 1% of non-*Prospera* beneficiary children and youth attend indigenous primary schools, whereas 10% of *Prospera* beneficiaries opt for such schools. At the secondary level, the majority of children from nonbeneficiary families are enrolled in general schools, whereas about 40% of beneficiary families are enrolled in telesecondary schools. In contrast, only 10% of non-*Prospera beneficiary* children/youth attend telesecondary school in grade 7.

The last column of Table 4 reports the cumulative fraction of dropouts at every grade during lower-secondary school. 20% of *Prospera*-beneficiary youth dropout prior to grade 9 in comparison to 16% of nonbeneficiaries. The higher dropout rates for the $P = 1$ group could, in part, explain the declining test-score gaps, if dropouts come disproportionately from the lower tails of the test-score distributions. For this reason, it is important to control for selective dropout in evaluating achievement test-score program impacts.

In Figure 2, we compare the test score distributions by grade and by lower-secondary-school types. The left column shows the distributions for the mathematics tests and the right column for the Spanish tests. Each row shows test scores for a different grade, from 6 to 9. There is a larger variance in mathematics test-score performance across school types than in Spanish performance. At grade 6, students who later enroll in telesecondary schools have a notably lower initial mathematics score. However, this gap begins to reverse by grade 7, ultimately resulting in substantially higher mathematics scores for *Prospera* students by grade 9. A comparison of the different school types shows that the general and technical schools consistently have higher Spanish scores than telesecondary schools at grade 6. However, this advantage diminishes as students progress through the grades. These graphs suggest that students attending telesecondary schools tend to demonstrate more significant improvements in test scores across

Table 3. Grade and enrollment distribution by age.

Age	G4	G5	G6	G7	G8	G9	Dropout	Obs.
9	1.00	–	–	–	–	–	–	35,346
10	0.78	0.22	–	–	–	–	–	157,931
11	0.10	0.71	0.20	–	–	–	–	172,709
12	0.03	0.10	0.68	0.19	–	–	0.01	176,581
13	–	0.03	0.10	0.65	0.18	–	0.05	175,984
14	–	–	0.03	0.09	0.61	0.17	0.10	176,237
15	–	–	–	0.03	0.10	0.70	0.17	141,657
16	–	–	–	0.01	0.15	0.43	0.41	19,026
17	–	–	–	–	0.04	0.42	0.54	3962

Note: The final column shows the number of observations for each age group. Excluding the last column, the sum of each row's values equals 1.

TABLE 4. Enrollment distribution by school type, grade, and *Prospera* status (P).

Primary school	General	Indigenous		
		Beneficiary ($P = 1$)		
Grade 4	0.89	0.11		
Grade 5	0.89	0.11		
Grade 6	0.89	0.11		
		Nonbeneficiary ($P = 0$)		
Grade 4	0.99	0.01		
Grade 5	0.99	0.01		
Grade 6	0.99	0.01		
Secondary school	**General**	**Telesecondary**	**Technical**	**Dropout**
		Beneficiary ($P = 1$)		
Grade 7	0.26	0.44	0.21	0.09
Grade 8	0.24	0.41	0.20	0.15
Grade 9	0.23	0.39	0.18	0.20
		Nonbeneficiary ($P = 0$)		
Grade 7	0.52	0.10	0.32	0.06
Grade 8	0.50	0.09	0.31	0.11
Grade 9	0.47	0.08	0.29	0.16

Note: This table displays the school-type enrollment distribution by beneficiary status and grades, where $P = 1$ denotes *Prospera* participation. The dropout only matters for lower-secondary school and is measured prior to entering that grade and is cumulative. Each row totals add to 1.

grades compared to students in other school types. However, these comparisons do not imply causation, as they do not account for the compositional differences among students in different school types and the different dropout rates across grades and school types. Our model will control for these different sources of selectivity.

Cheating behavior A potential concern in any assessment of student test scores is student cheating. Most studies ignore the issue of cheating, because researchers commonly lack access to information on who may have cheated. However, SEP employs a statistical algorithm designed to flag potential instances of cheating in the form of copying on the ENLACE exams. This information is integrated into our test-score database.[16]

Table 5 presents the impact of potential cheating behavior on test scores. When comparing students identified by the cheating indicator with those who did not cheat, average test scores in both mathematics and Spanish are higher for those flagged as cheating. The average test-score disparities are more pronounced in secondary schools, suggesting that gains from cheating might be greater for older children. A comparison between non-*Prospera* students and *Prospera* students reveals that cheating behavior is more

[16] In particular, SEP estimates a student's probability of cheating using both the K-index and the Scrutiny method. If the software detects a possible cheating case from the response patterns, a note is added to the student's report. Note that the exam score is not deleted, and there are no consequences to the students (SEP (2010)).

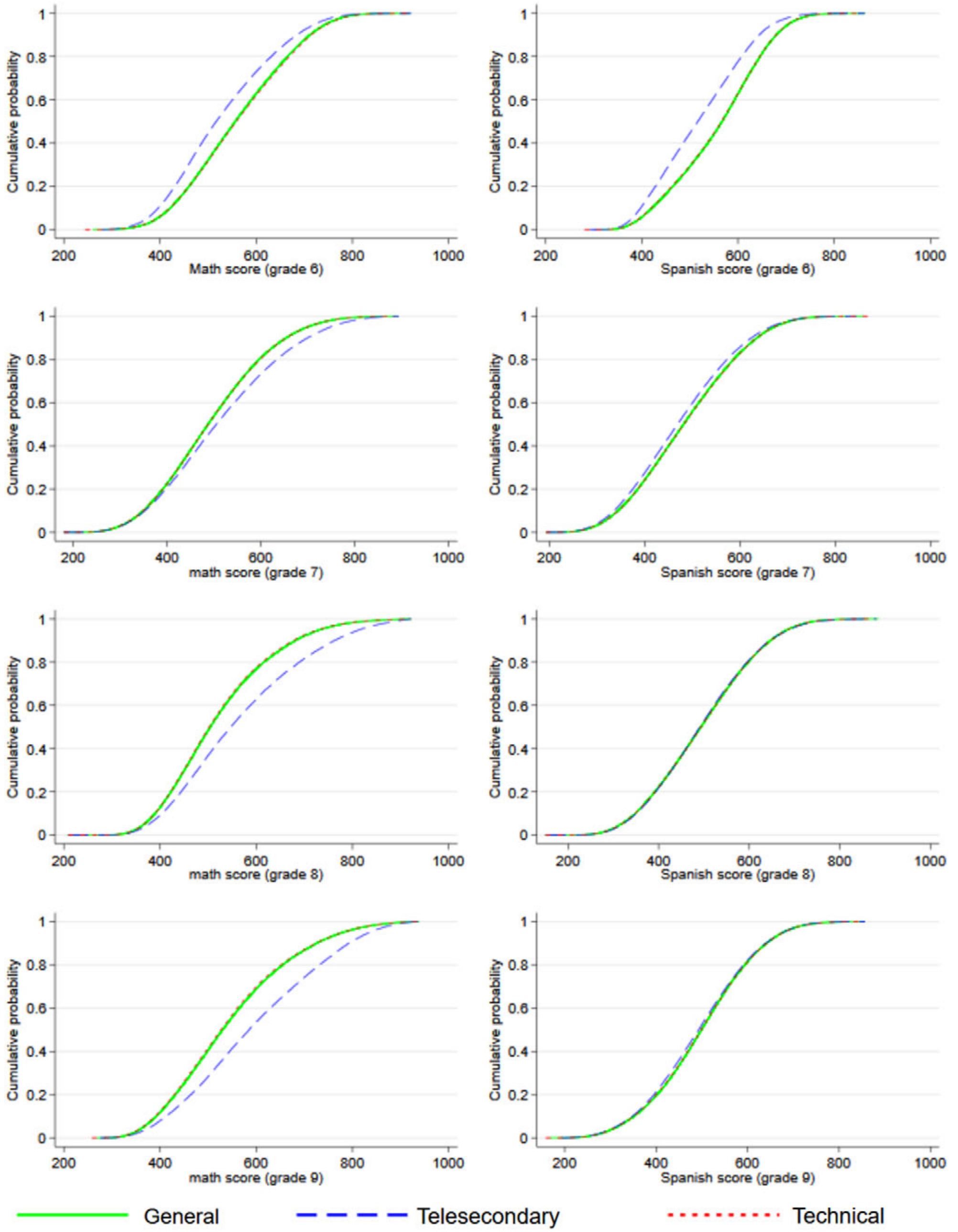

Figure 2. Mathematics and Spanish test scores distributions by grade and lower-secondary school types.

prevalent among the *Prospera* students. Moreover, cheating is generally associated with less test score inflation among non-*Prospera* than *Prospera* students.

TABLE 5. Average test scores and proportion cheating, by *Prospera* status.

Grades		Non-Prospera		Prospera	
		No cheating	Cheating	No cheating	Cheating
Grade 4	Fraction	93.8%	6.2%	90.1%	9.9%
	Math	529	554	480	519
	Spanish	521	538	468	496
Grade 5	Fraction	96.6%	3.4%	94.6%	5.4%
	Math	533	585	494	557
	Spanish	534	568	488	533
Grade 6	Fraction	96.9%	3.1%	95.1%	4.9%
	Math	560	616	525	594
	Spanish	557	590	515	561
Grade 7	Fraction	98.2%	1.8%	96.5%	3.5%
	Math	500	578	492	602
	Spanish	490	542	465	534
Grade 8	Fraction	96.2%	3.8%	92.7%	7.3%
	Math	524	627	526	683
	Spanish	497	567	476	579
Grade 9	Fraction	97.2%	2.8%	95.1%	4.9%
	Math	548	636	562	654
	Spanish	501	541	481	532

Note: This table shows average mathematics and Spanish test scores, along with the proportion of students flagged for cheating, disaggregated by *Prospera* status and grade. "Cheating" refers to students flagged by a copying behavior indicator.

2.5 *Local school supply and quality*

As previously noted, Mexican families can choose among different types of schools, but their options depend on the local supplies. We next examine the supplies of different types of schools and also their quality characteristics. Table 6 provides information on the supplies of local schools of different types at the individual level. In Mexico, multiple school sessions are often held in the same building, such as a morning and afternoon session. The different sessions may have different principals and teachers; so, in the data set they are considered to be different schools.[17] Primary schools tend to be small, with an average enrollment of less than 200 and, consequently, there are a large number of primary schools. Their small size partly reflects that the school systems do not usually provide transportation and students typically walk to school. Also, 77% of children do not have access to indigenous schools, which are typically located in areas with significant indigenous populations. At the lower-secondary level, there are fewer schools and they are larger.

Table 7 compares the different school types in terms of some average school quality characteristics, including pupil–teacher ratios and teacher educational levels. The

[17]In SA Figure B1, we show one illustrative example of local primary school sessions in Aguascalientes, a city in central Mexico. It has 316 school sessions distributed in 250 unique coordinates within 10 kilometers.

TABLE 6. Number of local schools of different types.

	Mean	Std	p10	p50	p75	Not available
			Primary school (within 5 km)			
General	63	78	9	25	100	3.4%
Indigenous	5	6	1	3	6	77.2%
			Secondary school (within 10 km)			
General	46	67	4	17	60	14.9%
Telesecondary	13	11	4	10	19	8.5%
Technical	10	11	2	6	15	16.1%

Note: Columns 3–5 report selected percentiles. The last column gives the percentages of individuals for whom a given school type are not locally available.

first two columns show characteristics for general and indigenous primary schools. Indigenous primary schools have on average 94 students in comparison to 174 students in general primary schools. The percentages of students who are disabled ranges from 1–2%. Despite having overall fewer students, the student–teacher ratio in indigenous schools is higher—33 in comparison to 24. Another difference is that teachers in indigenous schools are more likely to have only upper-secondary school degrees (17% in comparison to 3%). At the same time, the fraction of teachers with an undergraduate or higher degree is 7 percentage points higher. Thus, teacher schooling attainment exhibits higher variance in indigenous schools.

The last three columns of Table 7 compare the average school characteristics for general, technical, and telesecondary schools. Technical schools tend to be larger, with an average enrollment of 395 in comparison to 296 for general schools and 75 for telesecondary. Again, the proportion of disabled students across all types of schools is 1–2%. The student–teacher ratio is 14 in general schools, 19 in technical schools, and 24 in telesecondary schools.[18] Thus, we see a general pattern of the smaller schools in rural areas having higher student–teacher ratios, which could either reflect that video learning is less teacher-intensive or that teacher or resource shortages are more common in rural areas. Comparing teacher educational profiles across the different kinds of secondary schools, we see that average characteristics are fairly similar. The main difference is that general school teachers are more likely to have undergraduate degrees rather than teaching-college degrees, compared to teachers in technical and telesecondary schools.

3. MODEL

Our modeling framework combines a school-choice model of attendance decisions at different types of schools with models of academic achievement in mathematics and Spanish that are linked across ages/grades. In particular, we specify test-score gains from year-to-year using a value-added framework that relates current achievement to

[18]These tabulations are based on regular teachers and exclude art and music teachers who often teach at multiple schools.

Table 7. Mean primary and lower-secondary school characteristics by school types (with standard deviations in parentheses).

	Primary		Lower Secondary		
Characteristic	General	Indigenous	General	Telesecondary	Technical
Number of students	174	94	296	75	395
	(175)	(95)	(244)	(64)	(247)
Proportion disabled	0.02	0.01	0.01	0.01	0.02
	(0.06)	(0.07)	(0.04)	(0.03)	(0.05)
Student–teacher ratio	24	33	14	24	19
	(18)	(12)	(7)	(9)	(8)
Teachers with HS degree	0.03	0.17	0.03	0.03	0.02
	(0.09)	(0.30)	(0.07)	(0.16)	(0.06)
Teachers with teacher college	0.47	0.26	0.34	0.41	0.42
	(0.35)	(0.34)	(0.34)	(0.42)	(0.32)
Teachers with undergraduate degree	0.47	0.56	0.54	0.44	0.47
	(0.34)	(0.39)	(0.34)	(0.42)	(0.32)
Teachers with post-grad degree	0.03	0.01	0.07	0.11	0.08
	(0.10)	(0.07)	(0.12)	(0.23)	(0.11)

Note: Tabulations based on a school census data set called the 911 data.

lagged achievement, family and school inputs into the learning process, and student unobserved heterogeneity (e.g., arising from ability or preferences). Our framework also allows technology for producing test-score gains to vary by type of school. In addition, it incorporates drop-out decisions and allows for grade retention, as described below.

Our modeling framework can be considered quasistructural. The educational production function has a structural interpretation as a technology relating inputs to outputs. However, the school-choice model is reduced form, likely reflecting the decisions of students, parents, and school administrators. As discussed below, the outside option in the school-choice model is to drop out.

3.1 *General environment and sequential outcomes*

Individuals are indexed by i, $i = 1, \dots, n$ and each model period corresponds to one school year. In the initial period (a_f, corresponding to the age at grade 4), students/parents can choose to attend one of two types of primary schools: general ($j = 1$) or indigenous (bilingual) ($j = 4$), depending on the types locally available (within 5 km). At the end of grade 6, students simultaneously make school-enrollment decisions (with $j = 0$ indicating nonenrollment) and school-type choices from up to three options: general ($j = 1$), telesecondary ($j = 2$), or technical ($j = 3$), depending on the types locally

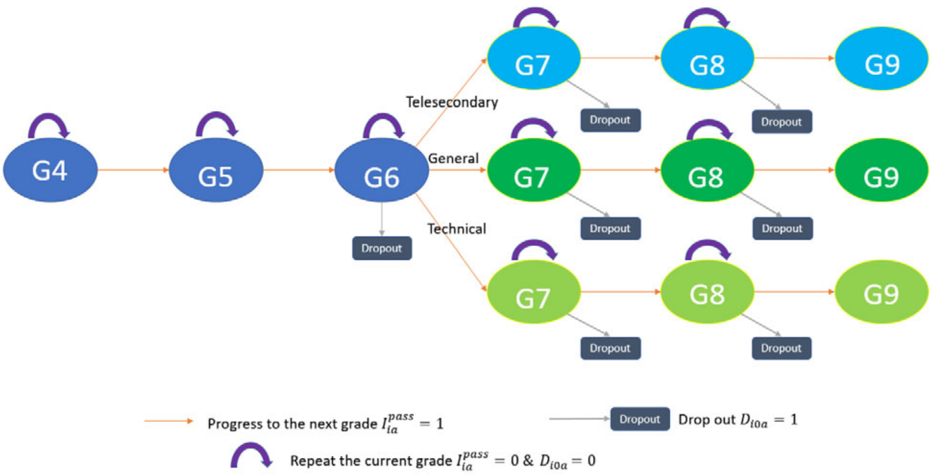

FIGURE 3. Potential sequential outcomes from grades 4 to 9.

available (within 10 km). We can summarize the choice set J_{ia}^{g} at different grades as

$$j_{ia} \in J_{ia}^{g} = \begin{cases} \{1, 4\} \cap M_i^1, & G_a = 4 \ \& \ a = a_f, \\ \{0, 1, 2, 3\} \cap M_i^2, & G_{a-1} = 6 \ \& \ I_{i,a-1}^{\text{Pass}} = 1, \\ \{0, j_{i,a-1}\}, & G_{a-1} \geq 7, \end{cases} \tag{1}$$

where a_f is the age when the student enters into the sample at grade 4. M_i^1 denotes the available local primary school types and M_i^2 denotes the available local lower-secondary school types.[19]

Let $D_{ija} = 1$ if the individual i is enrolled in school type j ($j \in \{1, 2, 3, 4\}$) at age a, else $D_{ija} = 0$. Let $D_{i0a} = 1$ if the individual does not enroll in school at age a, else $D_{i0a} = 0$. Let $I_{ia}^{\text{Pass}} = 1$ if the individual passes the grade in which she is enrolled at age a, else $I_{ia}^{\text{Pass}} = 0$. We assume the passing outcome I_{ia}^{Pass} is realized at the end of the school year, prior to the other decisions being made. The potential sequential outcomes from grade 4 to grade 9 are illustrated in Figure 3.

As seen in Figure 3, until grade 6 the possible outcomes are whether a student is retained in the current grade ($I_{ia}^{\text{Pass}} = 0$) or progresses to the next grade ($I_{ia}^{\text{Pass}} = 1$), conditioning on their primary school types when entering into the model at grade 4, their family background, and their *Prospera*-beneficiary status. Upon passing grade 6 ($I_{ia}^{\text{Pass}} = 1$), students simultaneously make enrollment decisions D_{i0a} and school choices D_{ija}, depending on locally available school types.[20] Once enrolled in a secondary school, students decide in each period whether to drop out or to stay in school. In each grade, there is a probability of passing or having to repeat the grade. Let G_{ia} denote the grade

[19] In particular, $M_i^1 \in \{\{1\}, \{4\}, \{1, 4\}\}$ and $M_i^2 \in \{\{1, 2, 3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1\}, \{2\}, \{3\}\}$.

[20] Because school enrollment is very high during primary school, we assume students do not drop-out during primary grades. Therefore, they do not make choices about continuing in school until the end of grade 6.

that the individual is eligible to attend at age a, which increases by one if the student passes the current grade:

$$G_{i,a+1} = G_{ia} + I_{ia}^{\text{Pass}}.$$

3.2 *Accounting for selective program participation*

Our aim is to use our estimated model to assess the impacts of *Prospera* participation on schooling progression and academic achievement, where we treat *Prospera* beneficiary status as a family characteristic. $P_i = 1$ denotes that a child/youth comes from a *Prospera*-beneficiary family, else $P_i = 0$. Although the survey data we use were not collected for the purpose of ascertaining Prospera eligibility, the data are rich and contain information on most of the eligibility determinants.[21] Program eligibility is not means-tested by income, because income can be difficult to measure in a country with a significant informal sector and where many low-income individuals are engaged in agricultural work. The program-eligibility criteria rather depend mainly on households' assets (such as car ownership), on characteristics of the household's residence (such as whether it has dirt floors, piped water, and how many rooms there are per person living in the house) and on household demographics, such as number of children and numbers of dependents per worker.[22] The vast majority of eligible households opt to participate, reflecting that the cash transfers are substantial.[23]

Our empirical strategy in evaluating the impact of *Prospera* on student test scores and educational progression is to first limit the comparison group subsample to children whose families meet at least a subset of the eligibility criteria. We do so by first estimating a probit model for the probability that each family is eligible for and participates in the *Prospera* program given the available information. That is, we use information on housing characteristics and demographics that were gathered through the student and parent surveys to estimate a household's probability of being eligible and participating in the program (a propensity score). The estimated coefficients from this probit regression are shown in SA Table A.3. The percentage correctly classified as being beneficiaries or not under the estimated model is high (90%).

Figure 4 plots the propensity-score distributions for children from *Prospera*-beneficiary households (in red) and nonbeneficiary households (in green). As seen in the figure, a large fraction of nonbeneficiaries fall in the first histogram bin, meaning that they have extremely low probabilities of participating in *Prospera*, generally because their characteristics make them ineligible. To increase comparability between the *Prospera* and the comparison-group subsamples, we impose a common support restriction and exclude in our impact analysis *Prospera* beneficiaries and the nonbeneficiaries with

[21] The precise eligibility criteria are not made public, but some of the authors of this paper were involved in the design of the criteria.

[22] Families who apply to the program typically fill out a questionnaire to determine their eligibility and their answers on the questionnaire may be checked through home visits.

[23] As discussed in Parker and Todd (2017), they represent on average about a 20% increase in household income.

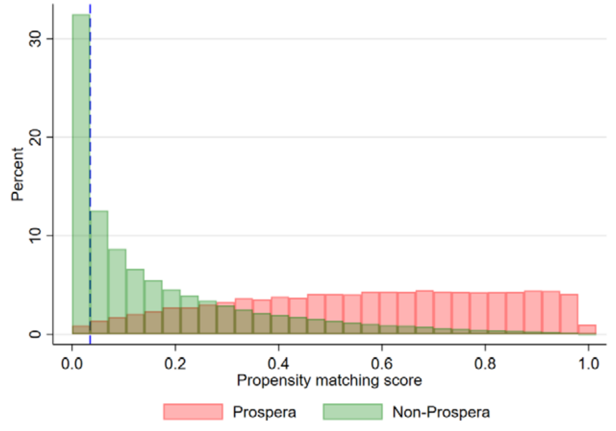

FIGURE 4. The propensity score distribution by *Prospera* status (P). *Note*: The red histogram represents the propensity score distribution for *Prospera* children/youth and the green histogram for *non-Prospera* children/youth.

propensity scores below the 1% quantile (the lowest bin of the histogram).This threshold excludes 383 children from *Prospera* families and 50,798 nonbeneficiary children.[24]

In addition, our school-choice/dropout and value-added models also include observed covariates to further control for differences between *Prospera* and non-*Prospera* households (such as parents' schooling attainment). We also allow for the possibility that children/youth from *Prospera* beneficiary families may differ in unobserved ways by including latent unobserved heterogeneity, specified as four discrete multinomial types that enter into all the model equations.[25] Let $\mu_{il} = 1$ denote that individual i is of type l, $=0$ else, where $l \in \{1, \dots, L\}$ and $L = 4$.

Ideally, one could allow for the unobserved types to be arbitrarily correlated with the observed variables. However, identifying such a model poses challenges, as it can be difficult to distinguish the direct effects of these observed variables from their indirect effects operating through the unobserved type distribution. For this reason, we restrict the type probability distribution to depend only on a child's *Prospera* status ($P \in \{0, 1\}$) and a binary marginality indicator ($M \in \{0, 1\}$), which is a measure of the poverty level in the locality where the household lives. We denote the conditional type probability as $\rho_l(P, M) \equiv \Pr(\mu_{il} = 1|P, M)$; it represents the fraction of type l among students with Prospera status P and marginality index M. Note that $\sum_l \rho_l(P, M) = 1$, $P \in \{0, 1\}$, $M \in \{0, 1\}$.

[24]This type of trimming is common in the application of matching estimators as a way of imposing "common support." Heckman, Ichimura, and Todd (1997) showed, in the context of evaluating a job-training program, that having a highly comparable comparison group is important to producing reliable nonexperimental impact estimates that replicate experimental estimates.

[25]See, for example, Heckman and Singer (1984), Cunha and Heckman (2008). Alternatively, we could impose a continuous distribution for the unobserved heterogeneity, for example, a mixture of normal distributions. Mroz (1999) shows the discrete-type assumption performs as well as the normal assumption when the true distribution is normal. When the true distribution is not normal, however, he finds that the discrete-type method performs better.

3.3 *The model*

As described in Section 2.4, our modeling and estimation approach is designed to address several statistical challenges that arise in evaluating the *Prospera* program academic achievement impacts. First, we address the problem of selective program participation by restricting our analysis sample to children estimated to have a positive probability of participating in the program.[26] Second, our school-attendance and school-choice framework, which models the sequential choices shown in Figure 3, explicitly addresses dynamic selection in school choices and dropout decisions. These decisions are permitted to depend both on observed family background characteristics as well as on unobserved factors that are assumed to follow a multinomial distribution (i.e., discrete types). We also incorporate exogenous variables, such as imputed local hourly wages, distances to the nearest school of each type, and the number of local schools of each type, as exclusion restrictions that affect school-type choices and dropout decisions but do not enter the test score outcome equations directly. Third, our model explicitly accounts for grade retention to capture that children may be observed multiple times in the same grade. Fourth, as described later in Section 4.1, we address potential test-score distortions caused by a small fraction of students suspected of cheating (copying).

We next describe our multiequation model of academic achievement over multiple grade levels, which includes the following components: value-added models in each grade, the school-choice/dropout models in primary school and at the start of secondary school, and the grade-repetition process.

Value-added model: Achievements in mathematics and Spanish evolve over time with school attendance. Let $m = 1$ denote mathematics, $m = 2$ Spanish, and g denotes the grade level. The value-added model is grade-specific and school-type (j) specific. The coefficients also vary depending on whether the student passed the previous grade $I^{\text{Pass}}_{i,a-1} = 1$ or is repeating the grade $I^{\text{Pass}}_{i,a-1} = 0$.[27] Let Z^A_{ia} denote the vector of observed characteristics of the youth and of the family that enter the achievement production function:

$$A^m_{ia} = \delta^{mgI}_{0jl} + A_{i,a-1}\delta^{gI}_{1j} + \delta^{mgI}_{2j} P_i + Z^A_{ia}\delta^{mgI}_{3j} + \omega^{mgI}_{ija}. \tag{2}$$

In this equation, δ^{mgI}_{0jl} is the type-specific intercept that allows for unobserved heterogeneity (l denotes the type). $A_{i,a-1} = \{A^1_{i,a-1}, A^2_{i,a-1}\}$ is a 2×1 vector including both the mathematics score and the Spanish score from the previous period $a - 1$. The lagged test-score terms are assumed to be sufficient statistics for the impacts of past inputs in the learning process. This specification allows for cross-effects between Spanish and mathematics. For example, better Spanish skills may enhance student's understanding in their mathematics classes, implying a positive effect of past Spanish scores on current mathematics scores. The impact of the *Prospera* program is captured by δ^{mgI}_{2j}. We

[26] The value of imposing common support on the propensity score distribution in the context of social program evaluation is emphasized in Heckman, Ichimura, and Todd (1997).

[27] If a student repeats a grade, then the lagged test score pertains to the same grade as in the current time period and would therefore have a different associated coefficient from the case where the lag pertains to the previous grade.

assume that the error terms ω_{ija}^{mgI}, conditional on the unobserved types, are i.i.d. and normally distributed. Most of the literature considers learning technology to be exogenous. By combining a school-choice model with value-added models that vary by school type, we allow students/parents to select from different available learning technologies.

School-choice model: We next specify how individuals make their schooling choice D_{ija} from the available options, J_{ia}^{g} (depending on his/her grade and geographic location, as defined in equation (1)). Assuming a random-utility model with Type I extreme-value errors (taste heterogeneity) yields a multinomial logistic model for the probability of choosing option j_{ia}:

$$\Pr\bigl(D_{ija}=1|\tilde{\Omega}(a),\mu_l\bigr) = \begin{cases} \dfrac{\exp\bigl(\mu_{0jl}^{g}+A_{ia-1}\phi_{1j}^{g}+P_i\phi_{2j}^{g}+f_j\bigl(Z_{ia}^{D},w_{ia},S_{ija}^{\text{distance}},S_{ija}^{\text{number}}\bigr)\bigr)}{\sum\limits_{j'\in J_a^g}\exp\bigl(\mu_{0j'l}^{g}+A_{ia-1}\phi_{1j'}^{g}+P_i\phi_{2j'}^{g}+f_{j'}\bigl(Z_{ia}^{D},w_{ia},S_{ij'a}^{\text{distance}},S_{ij'a}^{\text{number}}\bigr)\bigr)} & \text{if } j_{ia}\in J_a^g, \\ 0 & \text{if } j_{ia}\notin J_a^g, \end{cases} \tag{3}$$

where μ_{0jl}^{g} is a type-specific intercept (l denotes the type). The parameter ϕ_{1j}^{g} captures the effect of test scores on schooling choices, and ϕ_{2j}^{g} captures the impact of *Prospera* on schooling choices. $Z_{ia}^{D}\in\tilde{\Omega}(a)$ includes demographic and family-background characteristics. There are three additional variables that enter the school-choice equations but not other parts of the model: (i) the imputed hourly wage w_{ia}; (ii) the distance to the closest school of each type $S_{ija}^{\text{distance}}$; and (iii) the local supply of schools of each type S_{ija}^{number}.

As previously noted, the outside option in our school-choice model is to drop out of school, which could be a more attractive option in areas that pay higher wages to child labor. Due to the absence of wage information in our test-score databases, we rely on data from the 2010 Mexican census to impute wages for individuals based on characteristics such as their age, gender, schooling level, and geographical region of residence. Additionally, our imputation procedure incorporates a selection correction mechanism to account for selective labor-force participation.[28] The imputed wage, denoted as w_{ia}, represents the opportunity costs associated with being enrolled in school. Two other important variables in the school-choice model are the distance to the nearest school ($S_{ija}^{\text{distance}}$) and the log number of local schools of various types (S_{ija}^{number}). These variables are included to capture the effect of local school availability on individuals' schooling decisions.

We assume that the three exogenous variables $\{w_{ia},S_{ija}^{\text{distance}},S_{ija}^{\text{number}}\}$ affect school-choice decisions but do not directly enter the test-score equations (i.e., exclusion restrictions). These variables are helpful to identify separately the parameters in the value-

[28]For a more detailed explanation, please refer to the SA Table A.2.

added models from parameters in the school-choice/dropout model.[29] However, the exclusion restriction could be invalid, for example, if higher wages provide incentives for students to work part-time while enrolled in school, which directly affected their test-score performance. Another potential threat to validity is that the travel distance may directly affect the commuting time required to attend schools. Both channels may negatively impact academic performance through fatigue or reduced ability to concentrate on studying. To examine the empirical relevance of such concerns, we also estimated a specification in which the variables $\{w_{ia}, S_{ija}^{\text{distance}}, S_{ija}^{\text{number}}\}$ are added to the value-added equation (3). We examined (i) whether the coefficients associated with *Prospera*-beneficiary status change and (ii) whether the estimated coefficients associated with the variables $\{w_{ia}, S_{ija}^{\text{distance}}, S_{ija}^{\text{number}}\}$ are significantly different from 0. We did not find evidence of direct impacts of $\{w_{ia}, S_{ija}^{\text{distance}}, S_{ija}^{\text{number}}\}$ on test scores, conditional on the other model covariates.

As we will elucidate in Section 3.5, one way of interpreting our school-choice probability equation is an approximation to a policy function derived from a full dynamic version of our structural model (for related discussion see, e.g., Heckman, Humphries, and Veramendi (2016)). To account for potential nonlinear effects of the state variables on decision making, we adopt a flexible functional form, denoted as $f(Z_{ia}^{D}, w_{ia}, S_{ija}^{\text{distance}}, S_{ija}^{\text{number}})$.[30]

Grade retention: Lastly, we specify a probabilistic model for whether a student passes a grade, which depends on the unobserved type μ_l, the current school type attended j_{ia}, academic knowledge as proxied by the achievement scores A_{ia}, the grade level G_{ia}, *Prospera* beneficiary status P_i as well as some demographic and family-background characteristics, $Z_{ia}^{I} \in \tilde{\Omega}(a)$:

$$\Pr\big(I_{ia}^{\text{Pass}} = 1 | \tilde{\Omega}(a), \mu_l\big) = \Phi\big(\gamma_{0l}^{g} + A_{ia}\gamma_1^{g} + \gamma_2^{g} P_i + j_{ia}\gamma_3^{g} + Z_{ia}^{I}\gamma_4^{g}\big). \tag{4}$$

The coefficients of the passing probability probit model are grade-specific and γ_{0l}^{g} is a unobserved type-specific intercept.

3.4 *Treatment-effect heterogeneity*

Prospera may have heterogeneous impacts for students from different backgrounds. Inspired by the marginal-treatment-effect (MTE) literature, we divide the analysis sample into quartiles based on the *Prospera*-eligible propensity scores and allow the *Prospera* impact to vary by quartile.[31] Families in the highest quartile, that is, with characteristics

[29] In a parametric setting, exclusion restrictions are not strictly required. Nonetheless, independent variation in the determinants of dropout and school-choice decisions provides additional sources of identification that do not rely on functional form restrictions.

[30] In SA Section D, we describe how we use Bayesian Information Criterion (BIC) to determine our final econometric specification.

[31] We capture potential heterogeneity more parsimoniously than a standard approach in the literature, which is to estimate the treatment effect nonparametrically as a function of the propensity-matching score (e.g., Heckman and Vytlacil (2001, 2005, 2007)). However, most literature implements MTE in a static setup whereas our model is dynamic.

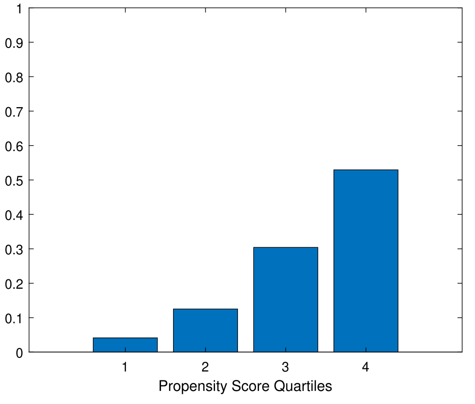

FIGURE 5. Distribution of *Prospera* students across propensity-score quartiles.

that make them most likely to be eligible for *Prospera* tend to be the most disadvantaged. Figure 5 shows the distribution of *Prospera*-beneficiary children across quartiles. A total of 83% of *Prospera* children/youth are concentrated in quartiles 3 and 4 (30% and 53%), reflecting the fact that the program is targeted at the poorest families.

Additionally, we also consider the possibility that the treatment effects vary by gender. Girls receive a higher conditional-cash-transfer subsidy than boys in secondary-school grades (see, e.g., Parker and Todd (2017)), so it is reasonable to explore whether *Prospera* impacts on academic achievement and attainment differ for girls and boys. We therefore allow, in estimation, for the coefficients associated with *Prospera* to vary by both gender and propensity-score quartiles. Specifically, we redefine the coefficients $\{\delta_{2j}^{mgI}, \phi_{2j}^{g}, \gamma_2^g\}$ to be $\{\delta_{2js\eta}^{mgI}, \phi_{2js\eta}^{g}, \gamma_{2s\eta}^{g}\}$, where $s = \{f, m\}$ denotes female and male, and $\eta = \{1, 2, 3, 4\}$ represents the quartile to which each individual belongs. As reported below, the most-disadvantaged children and youth experience substantially larger effects and there is evidence of heterogeneous impacts by gender.

3.5 *Approximate decision-rule interpretation*

As previously noted, the value-added achievement function has a structural interpretation as a production-function technology. Our school-choice model, shown in equation (3), can be interpreted as an approximation to a decision rule derived from a dynamic discrete-choice decision model. In a dynamic model, students' schooling decisions are age/grade-dependent and their educational choices in early grades affect their school attendance options in later grades.[32] The threshold-crossing indices associated with the school-choice probabilities at each age represent the differences between the value functions of alternative schooling-work options.

[32] Todd and Wolpin (2006), Attanasio, Meghir, and Santiago (2012) estimate structural schooling models to analyze the effects of the *PROGRESA* program on enrollment. Leite, Narayan, and Skoufias (2011) develop a model to evaluate impacts of the Brazilian *Bolsa Escola* CCT program. Those models consider school-enrollment/working decisions but do not consider the choice of school type or academic achievement. For discussion of how to approximate the decision rules see, for example, Keane, Todd, and Wolpin (2011), Heckman and Navarro (2007), Heckman, Humphries, and Veramendi (2018).

For example, consider students' choices at the end of grade 6 about whether to continue going to secondary school and, if so, which type of school to attend. The value function associated with each choice consists of the flow utility plus the expected future utility (the so-called Emax), which is typically a nonlinear function of the state variables (see equation (3)). We can approximate the value function using a flexible polynomial power-series. However, there are a large number of state variables and interaction terms that could potentially be included in such approximation. As described in SA Section D, we use a Bayesian Information Criterion (BIC) in selecting our final model specification, which approximates the value function of individual i choosing option j at age a as

$$\begin{aligned} V^*_{ija} = {} & \mu^g_{0jl} + A_{ia-1}\phi^g_{Aj} + P_i\phi^g_{pj} + Z^D_{ia}\phi^{1g}_{Zj} + w_{ia}\phi^{1g}_{wj} + S_{ija}\phi^{1g}_{Sj} + w^2_{ia}\phi^{2g}_{wj} + S^2_{ija}\phi^{2g}_{Sj} \\ & + \phi^{1g}_{Ij} \times \mathit{Region}_i \times \mathit{Urban}_i + \phi^{2g}_{Ij} \times \mathit{Region}_i \times S^{\text{distance}}_{ija} + \phi^{3g}_{Ij} \times \mathit{Region}_i \times S^{\text{number}}_{ija} \\ & + \phi^{4g}_{Ij} \times \mathit{Region}_i \times w_{ia} + \phi^{5g}_{Ij} \times \mathit{Urban}_i \times w_{ia} + \epsilon_{ija}. \end{aligned}$$

The coefficients $\{\phi^{1g}_{Ij}, \phi^{2g}_{Ij}, \phi^{3g}_{Ij}, \phi^{4g}_{Ij}, \phi^{5g}_{Ij}\}$ are associated with the nonlinear effects arising from the interaction terms among variables $\{\mathit{Region}_i, \mathit{Urban}_i, S^{\text{distance}}_{ija}, S^{\text{number}}_{ija}, w_{ia}\}$. We also considered interaction terms between Z^D_{ia} and other state variables, but these terms were not selected by the BIC criterion.

As previously noted, the variables $\{S^{\text{distance}}_{ija}, S^{\text{number}}_{ija}, w_{ia}\}$ constitute exclusion restrictions, which enter the school-enrollment/school-choice equations but not the test-score equations. Our specification allows the effects of the exclusion restrictions to vary across different regions and between urban and rural areas. For example, the same commuting distance may imply different commuting costs due to the availability of different local-transportation options. We include interaction terms to capture these geographic sources of heterogeneity.

There are a few advantages of adopting this kind of "quasistructural" approach, rather than estimating a fully structural dynamic model.[33] First, if multiple individuals are involved in the school-choice decisions (students, parents, school administrators), then our approach avoids having to model the decision processes of all of these agents and how they interact. Second, estimating a structural dynamic schooling model with academic achievement is complicated, because the state space would include two continuous lagged test scores, which usually necessitates the use of Emax approximation methods. Lastly, our model specification allows for considerable flexibility in how variables affect decision-making at different ages. In particular, it allows for heterogeneous impacts of local wages and school-type availability on school-enrollment and school-choice decisions.

A limitation of this type of quasistructural approach, however, is that the model does not distinguish between current and future utility components and, therefore, cannot be used to study effects of policies that might be implemented in the future, such as the effects of cash transfers given only upon completing certain future grades. This limits the range of counterfactual policies that can be considered. In the analysis reported below,

[33] See, also, related discussion in Heckman, Humphries, and Veramendi (2016).

we consider the effects of two time-invariant policy changes, eliminating *Prospera* and removing telesecondary schools from the choice set.

3.6 *Identification*

In our model for primary education, families make a one-time binary decision about what type of school the child attends, from up to two kinds of schools (general or indigenous). Many families do not live near indigenous schools, so those parents do not have any choice to make. Upon completing primary school, students/parents face a one-time multinomial choice: they can either drop out to work or enroll in one of three different types of schools (general, technical, or telesecondary). In grades 7 and 8, they only make the binary decision about whether to drop out or continue in the same type of school. Below, we discuss identification of the model parameters at these different stages. Our model is parametric and model parameters are estimated via maximum likelihood. Standard identification arguments for parametric models therefore apply, namely that the model parameters are identified as long as the Hessian matrix exists and is invertible (See, e.g., Amemiya (1985)). In our discussion below, though, we consider whether model parameters are identified under weaker functional form assumptions.

3.6.1 *Identification of model parameters pertaining to primary-school choices* Manski (1988) considers semiparametric identification of binary-choice index models in static and dynamic settings.[34] He showed that binary-choice index-model parameters are identified (up to scale) under an assumption that unobserved variables are fully independent of the regressors or under a weaker quantile (e.g., median) independence assumption.[35] His identification arguments apply to our primary school-choice model and to our probabilistic grade-retention equation.

At the end of grade 6, students make a decision about what type of secondary school to enter or whether to drop out, with the decision being observed for everyone. Our multinomial-logistic model with discrete unobserved types falls within the more general class of mixed logit models. Fox, il Kim, Ryan, and Bajari (2012) show nonparametric identification in such models.

3.6.2 *Identification of model parameters pertaining to lower-secondary school grades* Students face the decision to drop out at the end of 7th and 8th grades, with this choice only being visible for those who have not dropped out in earlier grades, which is essentially a dynamic selection process. Heckman and Navarro (2007) establish semiparametric identification of a class of dynamic discrete-choice models and they consider a model of optimal stopping time of schooling as their running example.[36] Their framework allows for general forms of duration dependence and unobserved heterogeneity. Let γ^t denote the parameters of the period t index equations and let F_ν^t denote the distribution of the period t unobserved shocks. As discussed in Heckman and Navarro (2007),

[34] Also, see Heckman, Humphries, and Veramendi (2016, 2018) for other applications and extensions.

[35] Honoré and Kyriazidou (2000) extend the identification results to a binary-choice panel-data model with lagged dependent variables and individual fixed effects.

[36] They call this model a time-to-treatment model, where the treatment is stopping school.

the set (γ^t, F_ν^t) is identified relative to all other sets of parameters $(\gamma^{*t}, F_\nu^{*t})$ if there exists a sequence of past choices such that the probability of observing a choice under the true parameters differs from that under alternative parameter values.[37] Their identification proof (Theorem 1) follows an identification-in-the-limit strategy that conditions on large values of the indices associated with preceding choice probabilities. Applying this identification argument in our setting, the dropout probability function at any grade can be identified using the subset of individuals who attain that grade level with probability close to 1. Although their proof does not require conventional exclusion restrictions, they note that the assumptions of their Theorem 1 will be satisfied if there are transition-specific exclusion restrictions (that satisfy certain rank and support conditions). In our context, distance to different types of schools, the number of schools available, and the hourly wage w_{ia} imputed from the census data (which varies with age, educational attainment, and region of residence) constitute exclusion restrictions.

3.6.3 *Identification of value-added model parameters* The value-added achievement model is a standard dynamic-panel-data model, of the kind discussed in Arellano and Bond (1991), with two modifications. First, the model incorporates discrete unobserved types to control for unobserved heterogeneity. Second, there is the potential for dropout after grades 6, 7, and 8, and the test score outcomes are only available for students who enroll in school. Heckman and Navarro (2007) also consider the identification of continuous outcomes and counterfactual outcomes in a stopping-time model, again using an identification-in-the-limit strategy based on sets of students with characteristics such that they advance to the next grade with a probability close to 1. Our permanent-transitory specification of the error term in the value-added equation is analogous to the one-factor error structure that they present. They note that one needs to have at least three outcome measures to identify a one-factor structure. In our case, we have 12 continuous outcomes (mathematics and Spanish test scores, from grade 4 through grade 9). They establish nonparametric identifiability of both the factor and error-term distributions, although we implement a parametric model.[38]

4. Estimation

4.1 *Some measurement issues*

In this section, we discuss two issues related to test-score measurement. First, some enrolled students are missing test-score data because they were absent the day of the test, perhaps due to illness. We use the notation $I_{ia}^{\text{miss}} = 1$ if the student's test score at age a is missing, else $I_{ia}^{\text{miss}} = 0$.

[37]Their theorem allows for general error structures F_ν^t that can be correlated over time for each individual but are assumed to be independent across individuals and of regressors in the initial time period. Our permanent-transitory error structure is a special case (see footnote 16 in Heckman and Navarro (2007)).

[38]Heckman and Navarro (2007) assume the factors are independent of covariates. We assume the type distribution can vary by *Prospera* status (an initial condition) and by the marginality index, which is a measure of local poverty. The Heckman and Navarro (2007) identification strategy can still be applied under this slightly more general formulation by first stratifying the data by *Prospera* status and marginality level.

Second, students' true test-score performances may be mismeasured if there is cheating. To account for possible test-score distortion due to cheating, we specify a measurement equation for the relationship between the true test score A_{ia}^m and the observed test score $\tilde{A}_{ia}^m$:

$$\tilde{A}_{ia}^m = \begin{cases} \left(1 + c_{ija}^{mgI} \zeta_{ia}^m I_{ia}^{\text{cheat}}\right) A_{ia}^m & \text{if } I_{ia}^{\text{miss}} = 0, \\ \text{Not observed} & \text{if } I_{ia}^{\text{miss}} = 1, \end{cases} \tag{5}$$

where the effect of cheating is captured by the term $c_{ija}^{mgI} \zeta_{ia}^m$. The parameter c_{ija}^{mgI} captures the average cheating effect for a given subject m, grade g and retention status I, while $\zeta_{ia}^m \sim \log\text{normal}(-0.5\sigma_\zeta^2, \sigma_\zeta^2)$ captures the randomness of the cheating effect.[39] When students do not cheat ($I_{ia}^{\text{cheat}} = 0$) and the test scores are not missing ($I_{ia}^{\text{miss}} = 0$), under this specification, the observed test score equals the true test score:

$$\tilde{A}_{ia}^m = A_{ia}^m \quad \text{iff} \quad I_{ia}^{\text{cheat}} = 0 \quad \text{and} \quad I_{ia}^{\text{miss}} = 0.$$

4.2 *The likelihood function*

The outcomes observed for individuals are the enrollment choices at each age D_{ija}, an indicator for whether the student passes the grade attended at age a I_{ia}^{Pass}, an indicator for whether the test scores are missing I_{ia}^{miss}, and when not missing, the achievement test scores in mathematics and Spanish at each age a, $\tilde{A}_{ia} = \{\tilde{A}_{ia}^1, \tilde{A}_{ia}^2\}$. Let a_i^f and a_i^l denote the first and the last ages at which we observe the individual i in the data set. The initial conditions include family background, achievement test scores at grade 4 (when test scores first become available for this sample), age, gender, a marginality index describing the local poverty level, urban/rural residence status, the state of residence, the set of primary and secondary schools available to each child, the minimum distances needed to travel to different types of schools and whether the family is participating in *Prospera*.[40] We use notation $\Omega_i(0)$ to denote the set of these initial state variables. The time-varying state-space elements in any given time period, $\Omega_i(a)$, consist of whether attended school last year, type of school attended, grades completed thus far, whether retained in the last grade, and lagged achievement-test scores in mathematics and Spanish.

Given the vector of the observed test scores $\tilde{A}_i = \{\tilde{A}_{ia_f}^1, \tilde{A}_{ia_f}^2, \ldots, \tilde{A}_{ia_l}^1, \tilde{A}_{ia_l}^2\}$, the vector of the true test scores $A_i = \{A_{ia_f}^1, A_{ia_f}^2, \ldots, A_{ia_l}^1, A_{ia_l}^2\}$, and the vector of initial conditions and time-varying state-space elements $\tilde{\Omega}_i = \{\Omega_i(0), \Omega_i(a_f), \ldots, \Omega_i(a_l)\}$, the indi-

[39] We choose this particular functional form so that $\zeta_{ia}^m > 0$ and $E(\zeta_{ia}^m) = 1$.

[40] A full list of family-background variables includes parental-schooling attainments, parental-employment statuses, whether both mother and father are present in the household, number of household members, first language spoken at home, internet accessibility, computer accessibility, and years of preschool education.

vidual likelihood can be written as

$$
\begin{aligned}
&L_i(\Theta, \mu, \rho; \tilde{A}_i, A_i, \tilde{\Omega}_i) \\
&= \sum_{l=1}^{4} \rho_l(P, M) \Bigg\{ \underbrace{\prod_{j \in J^4_{ia_f}} \Pr\big(D_{ija_f} = 1 | \Omega_i(0), \mu_l\big)^{1(D_{ija_f}=1)}}_{\text{Primary-school choice at initial period}} \\
&\times \prod_{a=a_f+1}^{a_l} \Bigg\{ \underbrace{\Pr\big(I^{\text{Pass}}_{i,a-1} = 1 | A_{i,a-1}, D_{ij,a-1}, \mu_l\big)}_{\text{Prob of passing the last grade } g_a} \prod_{j \in J^g_{ia}} \underbrace{\Pr(D_{i0a} = 1 | A_{ia}, \mu_l)^{1(D_{i0a}=1)}}_{\text{Prob of dropout at grade } g_a} \\
&\times \underbrace{\Big[\Pr(D_{ija} = 1 | A_{ia}, \mu_l) \phi\big(A_{ia} | A_{i,a-1}, D_{ij,a-1}, I^{\text{Pass}}_{i,a-1}, \mu_l\big) \phi\big(A_{ia} | \tilde{A}_{ia}, D_{ija}, I^{\text{Pass}}_{i,a-1}\big)^{1(I^{\text{miss}}_{i,a-1}=0)}\Big]^{1(D_{ija}=1)} \Bigg\}^{1(I^{\text{Pass}}_{i,a-1}=1)}}_{\text{Prob of observing test score } \tilde{A}_{ia} \text{ when passing grade } g_{a-1} \text{ in a school of type } j} \\
&\times \underbrace{\Big[\Pr\big(I^{\text{Pass}}_{i,a-1} = 0 | A_{i,a-1}, D_{ij,a-1}, \mu_l\big) \phi\big(A_{ia} | A_{i,a-1}, D_{ij,a-1}, I^{\text{Pass}}_{i,a-1}, \mu_l\big) \phi\big(A_{ia} | \tilde{A}_{ia}, D_{ij,a-1}, I^{\text{Pass}}_{i,a-1}\big)^{1(I^{\text{miss}}_{i,a-1}=0)}\Big]^{1(I^{\text{Pass}}_{i,a-1}=0)}}_{\text{Prob of observing test score } \tilde{A}_{ia} \text{ when repeating the grade in a type } j \text{ school}} \Bigg\}
\end{aligned}
$$

where Θ defines the vector of model parameters and we suppress the dependence of all the probabilities on the state space $\tilde{\Omega}_i$ to simplify notation. $\Pr(\cdot)$ represents the logit or multinomial-logit probabilities of school-choice and retention decisions defined in equations (3) and (4), while $\phi(\cdot)$ represents the conditional density function of test scores derived from equation (2) and equation (5). J^g_{ia} is the available school-choice set at grade g_a defined in equation (3). The vector $\rho = \{\rho_1, \dots, \rho_4 | P, M\}$ denotes the vector of unobserved-type probabilities conditioning on *Prospera* status P and marginality index M, where $\rho_l(P, M) = \Pr(\mu_l = 1 | P, M)$.

As described in the previous section, the true test scores are not observed in cases of cheating or when students are absent on the day of the test. In those cases, we integrate over the possible true test-score outcomes to obtain the individual likelihood function:[41]

$$L_i(\Theta, \mu, \rho; \tilde{A}_i, \tilde{\Omega}_i) = \int \dots \int L_i(\Theta, \mu, \rho; \tilde{A}_i, A_i, \tilde{\Omega}_i) dA_{ia^1_i} \dots dA_{ia^N_i},$$

where $\{a^1_i, a^2_i, \dots, a^N_i\}$ are the ages that the test scores are either contaminated by cheating or missing for individual i. We obtain standard errors of the parameter estimates from the inverse of the average of the product of the score matrices, where the derivatives of the log likelihood are evaluated numerically.[42]

4.3 *Some simplifying assumptions*

We impose three assumptions with regard to the choice problems that simplify the estimation and are largely consistent with the data. First, we assume students do not switch to a different type of school once enrolled. Therefore, the primary-school type is chosen

[41] This integration is performed numerically by taking 50 random draws of the shocks.

[42] This estimator is known as the BHHH estimator (Bernal and Keane (2010)). To obtain the numerical derivatives needed to implement the estimator, we use a step-size parameter equal to 1% of the parameter estimates.

once at the start of primary school and the lower-secondary school type is chosen once at the start of lower-secondary school. Second, we assume that individuals who drop out of school do not afterwards reenroll.[43] Third, because the fraction of students who repeat grades is fairly small (see Table 2), we estimate a separate value-added model for retained students but restrict the model coefficients to not vary across grades, separately within primary school and lower-secondary school grades.[44]

4.4 *Evaluating the effects of Prospera-program participation*

Prospera provides cash transfers for children of beneficiary households who are enrolled in school in grades 3–12. For children in grades 3–9, the transfers typically go to the mothers. Whether a transfer is received for each child depends, however, on whether that child regularly attends school (at least 85% of school days). We consider the family's *Prospera* status as a time-invariant characteristic, so it is contained in the initial state space $\Omega(0)$. Once families are enrolled in the program, they rarely lose their eligibility. Even if one child is not attending school, the family may still receive transfers for other children and is still considered to be participating. We use the estimated schooling model to simulate school-going and test-score outcomes for *Prospera* families' children had they not participated in *Prospera*. In this way, we are able to assess the grade-specific program impacts as well as the cumulative impacts of participating in *Prospera* for multiple years.

5. Model estimates

We estimate the model parameters by maximum likelihood. The key parameters, specifically those linked to *Prospera* effects, are shown in SA Section C, whereas the complete set of parameters can be found in SA Section D. In this section, we focus on two aspects: examining score distortion stemming from potential copying behavior, and evaluating the model's goodness-of-fit based on our adjusted test scores that account for test-score inflation resulting from cheating behavior.

5.1 *Test-score measurement equation*

A unique feature of our data set is that it contains information on which students were flagged by the SEP as potential copiers. Our test-score measurement equation allows the true test score to differ from the measured test score in the event of copying and also allows for heterogeneity in the gains from copying across grades and types of schools (to reflect potential differences in monitoring). In the context of a value-added model,

[43] In our raw data set, a mere 0.2% (378 individuals) switched schools to a different type during their primary-school years, while 3.2% (6220 individuals) changed school types in lower-secondary school. Additionally, we note that 1.99% (3770 individuals) reenrolled in school after initially leaving.

[44] That is, we restrict $\delta_{kj}^{m4I} = \delta_{kj}^{m5I}$, $\delta_{kj}^{m6I} = \delta_{kj}^{m7I} = \delta_{kj}^{m8I}$, $k = \{1, 2, 3\}$ in the value-added equation (2) and $\gamma_k^4 = \gamma_k^5 = \gamma_k^6$, $\gamma_k^7 = \gamma_k^8$, $k = \{1, 2, 3, 4\}$ in equation (4) in the periods when $I_{ia}^{\text{Pass}} = 0$. But the intercept terms δ_{0jl}^{mgI} and γ_{0l}^{g} are grade-specific.

TABLE 8. Estimated test-score distortion from copying, by grade and school type.

		Math			Spanish		
	Percentages	Raw	True	Diff	Raw	True	Diff
			Grade 5				
General	4.2%	570	525	45	549	521	28
Indigenous	7.0%	540	474	66	515	468	46
Overall	4.4%	568	521	47	546	517	29
			Grade 6				
General	4.0%	606	554	52	578	545	33
Indigenous	6.3%	566	510	56	536	499	37
Overall	4.1%	603	551	52	575	542	33
			Grade 7				
General	1.7%	558	499	59	523	489	34
Telesecondary	4.1%	619	525	94	540	486	54
Technical	2.5%	573	498	75	536	489	47
Overall	2.6%	588	510	78	534	488	46
			Grade 8				
General	3.3%	614	529	85	555	508	47
Telesecondary	9.9%	698	580	118	585	517	68
Technical	4.1%	613	528	85	554	511	43
Overall	5.3%	656	554	102	570	513	57
			Grade 9				
General	2.6%	631	549	82	531	505	26
Telesecondary	5.7%	667	607	60	529	509	20
Technical	3.8%	621	553	68	536	510	26
Overall	3.8%	642	574	68	532	508	24

Note: The percentages in the second column give the percentages of students suspected of copying in either mathematics or Spanish tests.

copying can lead to a one-sided measurement error in either the dependent variable (the test score) or in an independent variable (lagged test scores) or in both variables, but only for students who copied. Table 8 shows the percentages of students suspected of copying, which ranges from a low of 1.7% in 7th grade in general schools to a high of 9.9% in 8th grade in telesecondary schools. At the primary-school level, indigenous schools exhibit higher copying rates. We estimate that copying distorts average test scores by 20–118 points for copiers. Our earlier-described estimation approach accounts for this potential distortion.

5.2 *Model goodness-of-fit*

Our model involves a substantial number of parameters, in part due to our deliberate choice not to impose constraints in the value-added model coefficients across various grade levels. These parameters are mostly precisely estimated due to our large sample sizes. Tables 9 and 10 provide evidence on the model's goodness-of-fit. In Table 9, we compare average test scores across grades and by *Prospera* beneficiary status in the

TABLE 9. Goodness-of-fit for average test scores by *Prospera* status (P).

	Mathematics score				Spanish score			
	$P = 0$		$P = 1$		$P = 0$		$P = 1$	
Prospera	Data	Sim	Data	Sim	Data	Sim	Data	Sim
Grade 5	521	522	495	496	520	519	490	490
Grade 6	550	551	526	527	545	543	515	517
Grade 7	489	489	492	491	477	477	465	465
Grade 8	516	515	529	526	487	487	480	480
Grade 9	540	539	564	563	490	489	481	481

Note: We simulate the test scores 100 times for each individual. In this table, we adjust for any copying in both the simulation and the data.

data and based on model simulations. The data averages closely align with the model-simulated averages, with few exceptions. The estimated model reproduces the observed pattern where *Prospera* beneficiaries have lower average test scores in primary grades for both subjects. It also captures the trend of test-score disparities between beneficiaries and nonbeneficiaries diminishing in lower-secondary grades, and even reversing in mathematics.

Table 10 shows how well our model fits the school-type distribution. The model's predicted proportions closely match the data, with differences of no more than 0.02. Our model effectively captures two important data features: the higher probability of *Prospera*-beneficiary children attending telesecondary lower-secondary schools and their higher rates of dropping out. Additionally, it reproduces the observed dropout patterns across different grades.

6. Assessing cumulative *Prospera*-program effects

6.1 *The cumulative Prospera-program effects*

Average treatment effect on treated As previously described, educational production functions typically assume that knowledge acquisition in mathematics and Spanish is

TABLE 10. Goodness-of-fit to school-type distribution.

Lower-secondary choice	General		Telesecondary		Technical		Dropout	
	Data	Sim	Data	Sim	Data	Sim	Data	Sim
				Nonbeneficiary ($P = 0$)				
Grade 7	0.49	0.47	0.13	0.13	0.31	0.30	0.08	0.09
Grade 8	0.46	0.44	0.12	0.12	0.29	0.28	0.13	0.15
Grade 9	0.41	0.40	0.11	0.11	0.26	0.26	0.23	0.23
				Prospera beneficiary ($P = 1$)				
Grade 7	0.26	0.24	0.43	0.44	0.21	0.22	0.10	0.10
Grade 8	0.24	0.23	0.41	0.41	0.20	0.20	0.16	0.16
Grade 9	0.21	0.20	0.36	0.36	0.17	0.18	0.26	0.26

Note: We simulate the test scores 100 times for each individual.

a cumulative process. The value-added model specification allows lagged knowledge to have an effect on contemporaneous knowledge accumulation, so that the history of inputs into the learning process matters. If *Prospera* participation increases knowledge at a particular grade, then this benefit can have a persistent effect on learning in future grades. That is, program participation can have both direct effects on current test scores as well as indirect effects operating through lagged test scores.

In Table 11, we use our estimated model to simulate the effects of being a *Prospera* beneficiary over multiple grades, starting with grade 4. Our estimation procedure allows for *Prospera* effects that operate through all of the different channels of our school progression and achievement model and that may differ for girls and boys. Columns labeled $P = 1$ show the outcomes for *Prospera*-beneficiary children/youth with their participation in the program. Columns labeled $\tilde{P} = 0$ show the simulated (counterfactual) outcomes were they not to participate in the program.[45]

It is only possible to assess test-score impacts for children/youth who would attend school both with and without *Prospera*. Therefore, our reported program impacts on test scores in column "Diff" represent lower bounds, as they do not include potential academic achievement gains for children/youth who in the absence of *Prospera* would not be attending the grade.[46] Our results show positive benefits of being a *Prospera* beneficiary in lower-secondary grades but essentially no effect for math and Spanish in primary grades. In lower-secondary school, the cumulative *Prospera* impact in mathematics increases with the grade level and reaches a high of 0.21 standard deviations by grade

Table 11. Cumulative program impacts.

	Mathematics score				Spanish score			
	$P=1$	$\tilde{P}=0$	Diff	S.E.	$P=1$	$\tilde{P}=0$	Diff	S.E.
Grade 5	495	496	−0.2	0.7	490	491	−0.9	0.7
Grade 6	527	522	4.3	0.8	517	521	−4.1	0.7
Grade 7	492	478	13.7	2.0	465	459	6.3	1.6
Grade 8	527	513	14.2	1.8	481	476	4.8	1.5
Grade 9	562	541	20.8	2.2	482	478	4.0	1.8
	Dropout rate				Retention rate			
	$P=1$	$\tilde{P}=0$	Diff	S.E.	$P=1$	$\tilde{P}=0$	Diff	S.E.
Grade 5	–	–	–	–	0.04	0.04	−0.004	0.002
Grade 6	–	–	–	–	0.02	0.03	−0.003	0.001
Grade 7	0.11	0.17	−0.07	0.01	0.003	0.002	0.000	0.000
Grade 8	0.16	0.23	−0.08	0.01	0.003	0.004	−0.001	0.001
Grade 9	0.25	0.33	−0.08	0.01	0.003	0.004	−0.001	0.001

Note: We report test score impacts for children/youth who would attend school both with and without *Prospera*. The cumulative impacts are obtained through bootstrap simulation with 100 replications. In particular, we first draw the model parameters from their estimated distributions and simulate the cumulative test scores and impacts for each bootstrap iteration. Then we obtain standard errors from the empirical distributions. The columns "Diff" capture the test-score gain of these subgroups. The columns "S.E." report the standard errors of the test-score gains from the program.

[45] Our simulation keeps the distribution of unobserved types for *Prospera* beneficiaries fixed.

[46] As we show in Figure 6 and Figure 7, absent children are disproportionally from most disadvantaged family backgrounds and, therefore, tend to have larger program gains on average.

9. In Spanish, the cumulative gains are substantially smaller — about 0.04 standard deviations.[47]

We might expect the estimated program effects to be larger in lower-secondary school than primary school because the transfer amounts that families receive for school attendance are much larger.[48] Also, older children typically have more demands on their time that compete with schoolwork than do younger children, such as taking care of younger siblings, housework, working for family businesses, or working for pay after school. The *Prospera* cash transfers may reduce these outside time uses, allowing them to focus more on schoolwork.

The lower panel of Table 11 reports the *Prospera* impact on the dropout rate (cumulative) and on the probability of repeating a grade. The results show that *Prospera* reduces the dropout rate by 0.08 before the start of grade 9, with the most pronounced effect during the primary to lower-secondary school transition. We also find that *Prospera* primarily reduces the retention probability during primary school but has no significant effect on retention during lower-secondary school.

Comparison of program effects for female and male students The *Prospera* program provides greater subsidies to girls than boys for attending school in post-primary grades. Table 12 examines whether program effects differ by gender. The test-score impacts are significantly positive for both girls and boys in all grades except for grade 5. The largest impacts are observed in lower-secondary school grades and impacts are larger in mathematics than in Spanish.

Interestingly, the estimates indicate that participation in *Prospera* leads to a slight reduction in gender academic achievement gaps. Male students have a 7-point advantage in average mathematics scores over females at grade 6. Participation in the *Prospera* program through grade 9 boosts female students' mathematics scores by 21.8 points in comparison to 19.6 for males and reduces the gender gap by 3.0 points (=(564-560)-(542-541)). In terms of Spanish test scores, female students have on average a 29-point advantage over males at grade 6. *Prospera* participation is also associated with a reduction in the gender gap in Spanish scores by grade 9, albeit by a slightly smaller margin of 2.0 points (=(495-467)-(492-462)).

The program also narrows the gender gap in dropout rates. For the current *Prospera* beneficiaries, the cumulative dropout rate by grade 9 is 23.3 percentage points for females and 27.0 percentage points for males. When we simulate the model taking away the *Prospera* program, we find that the gender difference increases from 3.7 (=27.0-23.3) percentage points to 6.3 (=36.4-30.1) percentage points. Turning to the bottom panel of the table, we do not find significant gender differences in the *Prospera* effect on grade retention. In summary, our results show that both females and males benefit from *Prospera* participation and that the program generally reduces gender disparities in mathematics and Spanish test scores and in dropout rates.

[47]The average effect on the Spanish score displays substantial heterogeneity among *Prospera* beneficiaries, as will be shown below.

[48]In the fall semester of 2008, the transfers ranged from 130 to 265 pesos for primary school and 405 to 495 (385 to 430) for females (males) in lower-secondary school (US1= 11 pesos in 2008).

TABLE 12. Gender differences in cumulative *Prospera* effects (by grade 9).

	Female				Male			
	$P=1$	$\tilde{P}=0$	Diff	S.E.	$P=1$	$\tilde{P}=0$	Diff	S.E.
				Mathematics score				
Grade 5	500	499	0.7	0.9	491	492	−1.1	1.1
Grade 6	530	525	4.7	1.1	523	519	3.9	1.2
Grade 7	495	481	14.2	2.4	488	475	13.3	2.2
Grade 8	525	511	14.9	2.0	530	516	13.5	2.0
Grade 9	564	542	21.8	2.5	560	541	19.6	2.5
				Spanish score				
Grade 5	503	505	−1.3	0.9	476	476	−0.5	1.0
Grade 6	531	536	−4.5	0.9	502	506	−3.6	1.0
Grade 7	484	480	4.6	1.9	446	437	8.1	1.7
Grade 8	498	494	4.2	1.8	462	456	5.5	1.7
Grade 9	495	492	3.2	1.9	467	462	5.0	2.1
				Dropout rate				
Grade 7	0.10	0.16	−0.06	0.01	0.11	0.18	−0.07	0.012
Grade 8	0.15	0.22	−0.07	0.01	0.16	0.25	−0.09	0.011
Grade 9	0.23	0.30	−0.07	0.009	0.27	0.36	−0.09	0.009
				Retention rate				
Grade 4	0.03	0.03	−0.003	0.002	0.05	0.06	−0.004	0.003
Grade 5	0.01	0.02	−0.002	0.002	0.03	0.04	−0.003	0.002
Grade 6	0.001	0.001	0.000	0.000	0.004	0.004	0.000	0.000
Grade 7	0.001	0.002	−0.001	0.001	0.005	0.006	−0.001	0.002
Grade 8	0.002	0.002	−0.001	0.001	0.005	0.007	−0.002	0.001

Note: Estimates obtained through model simulation. See the note to Table 11.

Prospera effects by propensity-scores quartiles We next explore the heterogeneous *Prospera* impacts for students from different backgrounds in Figures 6 and 7. Figure 6 shows the effects of *Prospera* participation on test scores broken down by propensity-scores quartiles. As described in Section 3.4, the propensity score is a summary statistic of students' family background, with quartile 1 denoting the most advantaged families and quartile 4 denoting the most disadvantaged families. Our estimates show larger impacts in later grades and smaller impacts in earlier grades, regardless of propensity-score quartiles. These patterns are consistent with the *Prospera* effects being cumulative with greater exposure associated with greater impact. Among the four quartiles, we observe the largest estimated cumulative impacts for students in the highest quartile, who are the ones from the most disadvantaged backgrounds. *Prospera* increases their test scores in mathematics by 0.29 standard deviations and their test scores in Spanish by 0.09 standard deviations and both effects are statistically significant. The top propensity score quartile contains the majority (52.7%) of the *Prospera* beneficiaries.

Figure 7 displays the cumulative effects of *Prospera* on three key outcomes: (i) Panel (a) presents grade attainment; (ii) Panel (b) presents the cumulative dropout rate; and (iii) Panel (c) presents the total number of retentions during primary school. The first two outcomes are evaluated at the end of a 6-year period, corresponding to the end

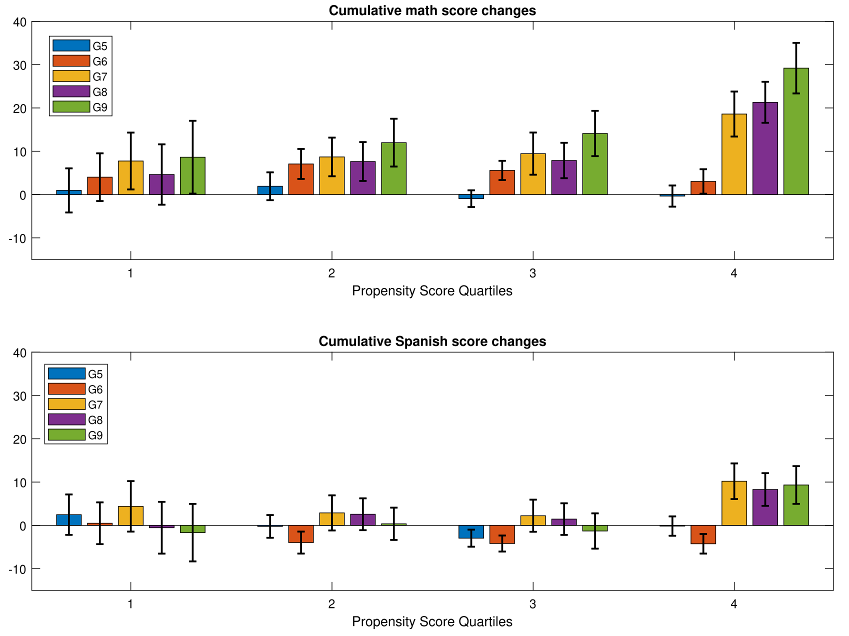

FIGURE 6. *Prospera* academic achievement effects by propensity-score quartiles. *Note*: 95% confidence intervals, depicted as bars, are derived using a parametric bootstrap method with 100 replications.

of grade 9 (for students who do not experience grade retention or drop out). The third outcome, the cumulative number of retentions, is assessed upon completion of primary school. The estimated impacts are shown conditional on the propensity score quartile. Notably, the most substantial impacts (with the exception of cumulative retentions) are

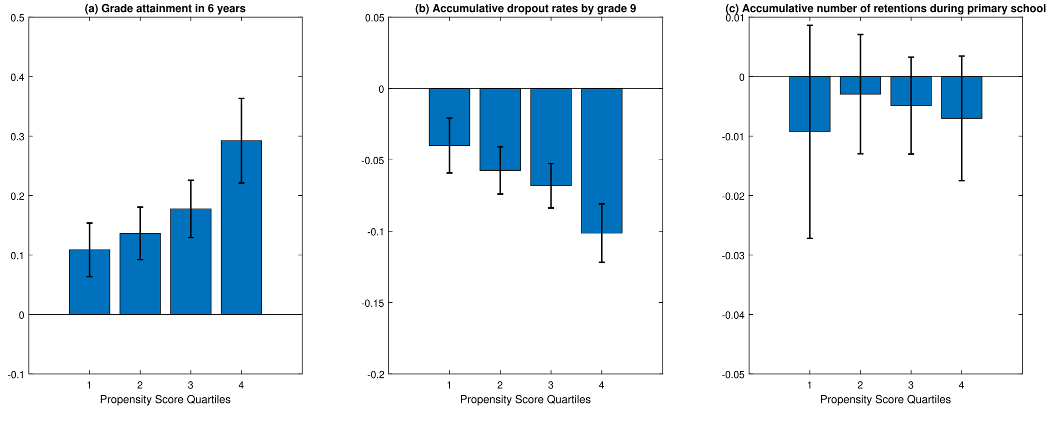

FIGURE 7. *Prospera* effects on schooling grade attainment, dropout, and number of retentions by propensity-score quartiles. *Note*: 95% confidence intervals, depicted as bars, are derived using a parametric bootstrap method with 100 replications.

observed in the fourth quartile, comprised of the most disadvantaged students. Because students opting to drop out are generally lower performing, disregarding the selection bias associated with dynamic dropouts tends to exaggerate the downward bias when estimating the program effects, particularly for students in the fourth quartile compared to those in other quartiles.

Comparing ATT with ATU Next, we compare the average treatment effect on individuals who received treatment, commonly referred to as "ATT," with the average treatment effect on those who did not receive treatment, commonly referred to as "ATU." In our analysis, the "untreated" group consists of children/youth who are not beneficiaries of the *Prospera* program ($P = 0$) but who had a positive probability of being a beneficiary, as determined by their household characteristics. (Recall that we imposed common support as described previously.)

The results in Table 13 indicate that the impact of the *Prospera* program is significantly greater for the treated group compared to the untreated group. For instance, the *Prospera* program leads to substantial improvements in mathematics and Spanish scores for the treated students, with increases of 20.8 and 4.0. For the untreated group, the program only results in a mathematics score improvement of 12.3 and has little effect on Spanish scores. At the extensive margin, *Prospera* reduces the dropout rate by 0.08 for the treated group but by only 0.05 for the untreated group. Lastly, *Prospera* also has a greater impact on reducing retention probabilities for the treated group compared to the untreated group.

These differences are in line with our earlier findings, shown in Figure 6. Those figures showed that the most substantial impacts are observed among the most disadvantaged groups. Now considering that these highly disadvantaged students are disproportionately more likely to be program beneficiaries, they predominantly fall into the treated group rather than the untreated group. Thus, it is expected that the ATT would be higher than the ATU.

6.2 *The importance of the telesecondary-school option*

As previously described, children/youth from *Prospera*-beneficiary households often live in rural areas where telesecondary schools are available and they more often attend this school type. We next evaluate the importance of telesecondary school as a determinant of *Prospera* impacts on school enrollment. In particular, we use our estimated model to simulate what educational outcomes would look like were the telesecondary-school option not available. The simulation takes into account that students might then have to travel further distances to get to schools or dropout if telesecondary schools had been their only option. Table 14 shows the distribution of local lower-secondary school-choice sets in the data (baseline) and after removing the telesecondary option. For 6.9% of students, telesecondary schools are the only option.

The upper panel of Table 15 shows the simulated dropout proportion (at grade 9) for current *Prospera* telesecondary enrollees when the telesecondary schools are removed from their choice sets. The dropout proportion increases dramatically from 0.18 to 0.52.

Table 13. Comparing ATT, ATU, and overall treatment effect.

	ATT		ATU		Overall	
	Mean	S.E.	Mean	S.E.	Mean	S.E.
			Mathematics score			
Grade 5	−0.2	0.7	0.8	1.1	0.4	0.8
Grade 6	4.3	0.8	5.2	1.1	4.9	0.9
Grade 7	13.7	2.0	9.1	1.8	10.7	1.7
Grade 8	14.2	1.8	7.6	1.9	9.9	1.7
Grade 9	20.8	2.2	12.3	2.4	15.2	2.1
			Spanish score			
Grade 5	−0.9	0.7	0.1	1.0	−0.3	0.8
Grade 6	−4.1	0.7	−2.5	1.0	−3.1	0.8
Grade 7	6.3	1.6	3.5	1.6	4.5	1.5
Grade 8	4.8	1.5	1.6	1.7	2.7	1.5
Grade 9	4.0	1.8	−0.2	1.8	1.2	1.7
			Dropout rate			
Grade 7	0.07	0.01	0.03	0.007	0.05	0.01
Grade 8	0.08	0.01	0.04	0.007	0.06	0.01
Grade 9	0.08	0.01	0.05	0.007	0.06	0.01
			Retention rate			
Grade 4	−0.004	0.002	−0.002	0.002	−0.001	0.001
Grade 5	−0.0025	0.0015	−0.0016	0.0011	−0.0009	0.0006
Grade 6	0.0001	0.0002	−0.0001	0.0002	0.0001	0.0001
Grade 7	−0.0009	0.0011	−0.0006	0.0008	−0.0003	0.0003
Grade 8	−0.0010	0.0008	−0.0005	0.0007	−0.0003	0.0002

Note: "ATT" denotes the average treatment effect on individuals who received the treatment, while "ATU" denotes the average treatment effect on those who did not receive it. The "untreated" group consists of children and youth who are not *Prospera* beneficiaries ($P = 0$) but had a positive probability of eligibility, based on household characteristics. Estimates are derived from model simulations.

Average educational attainment over the 6 years of our observation period (up to grade 9) falls from 8.76 grades to 7.57 grades. Despite using different data sources and evaluation approaches, our results align with evidence on telesecondary schools' importance reported in Navarro-Sola (2019).[49]

The lower panel of Table 15 shows simulated academic achievement for current *Prospera* telesecondary enrollees who continue their education even after telesecondary schools are no longer available. At grade 9, we observe a decrease in average mathematics test scores, from 602 to 535, and a decrease in average Spanish test scores from 492 to 473. This finding is consistent with the test score distributional differences seen in Figure 2, which suggested that telesecondary schools are relatively effective in enhancing students' test scores, particularly in mathematics,

[49]Using a difference-in-difference approach and Employment and Occupation National Survey (EONS) data set, she showed that the construction of an additional telesecondary per 50 children would encourage 10 individuals to enroll in lower-secondary education, causing an average increase of one additional grade of education among individuals that could have attended it.

TABLE 14. The school-choice distribution with and without the telesecondary option.

	Baseline	No telesecondary
General, technical and telesecondary	0.696	N/A
General and telesecondary	0.072	N/A
General and technical	0.061	0.757
Telesecondary and technical	0.078	N/A
Only general	0.012	0.090
Only telesecondary	0.069	N/A
Only technical	0.008	0.080
No local schools	0.004	0.073

6.3 *Quantifying the importance of dynamic selection*

The multiequation modeling framework that we implemented controlled for multiple sources of dynamic selection—due to dropout, school choice, and grade retention—as well as for cheating and missing data. It incorporated unobserved types to control for potential selectivity on unobserved factors. Arguably, in Mexico, selection is an important consideration, given that school enrollment drops significantly in lower-secondary school grades and that parents can select from available schools.[50] In the US context, value-added models are often implemented without accounting for selection.

To explore the importance of controlling for multiple sources of selection, we compare our baseline results with results obtained from a simpler value-added model that we estimate grade-by-grade without distinguishing school types:

$$A^m_{ia} = \delta^{mg}_0 + A_{i,a-1}\delta^g_1 + \delta^{mg}_2 P_i + Z^A_{ia}\delta^{mg}_3 + \omega^{mg}_{ia}.$$

Compared with equation (3), the contemporaneous *Prospera* effect δ^{mg}_2 is homogeneous across school types and we do not model school choice. Also, this model does not include permanent unobserved heterogeneity (types). The cumulative program effect can

TABLE 15. Simulated dropout, educational attainment, and achievement for *Prospera* telesecondary enrollees when the telesecondary option is removed.

	With telesecondary	Without telesecondary
Dropout rate	0.18	0.52
Grades attained	8.76	7.57
	Test scores at grade 9	
Math score	602	535
Spanish score	492	473

Note: The simulation is based on the *Prospera* beneficiaries who are currently enrolled in telesecondary school at grade 7. The outcomes are measured by grade 9.

[50] Cameron and Heckman (2001) consider the problem of selection in modeling grade progression in US high schools, but they do not analyze test-score data.

TABLE 16. Cumulative program impacts.

	Math score		Spanish score	
	(1) Baseline	(2) Simple VA	(3) Baseline	(4) Simple VA
Grade 5	−0.2	−0.4	−0.9	−4.0
Grade 6	4.3	−1.8	−4.1	−6.7
Grade 7	13.7	6.6	6.3	−1.2
Grade 8	14.2	10.4	4.8	0.6
Grade 9	20.8	15.4	4.0	0.8

Note: Simple VA = simple value added model.

be calculated as

$$\Delta_g^m = \begin{cases} \delta_2^{m5} & \text{if } g = 5, \\ \delta_2^{mg} + \delta_1^g \Delta_{g-1} & \text{if } g > 5, \end{cases}$$

where Δ_g^m represents the cumulative effect for subject m in grade g, $\Delta_{g-1} = [\Delta_{g-1}^1, \Delta_{g-1}^2]$ is a 2×1 vector of cumulative effects for both mathematics and Spanish in grade $g-1$.

Table 16 compares our baseline results for mathematics and Spanish test scores (columns (1) and (3)) with results generated from the simpler model (columns (2) and (4)). The cumulative program impacts derived from the simpler model are noticeably smaller, especially in the lower-secondary grades. As previously noted, not controlling for dropping-out will tend to downward bias the impact estimates if the program causes students at the margin of dropping-out to stay in school longer. Also, our previous analysis showed that *Prospera* beneficiaries attend telesecondary schools in greater numbers and that these schools were particularly effective in teaching mathematics. These benefits are not captured in the simpler model. Lastly, the simpler model also ignores the negative selection on unobserved types, which leads to underestimation of program impacts. In summary, we find that the richer modeling framework is needed to capture heterogeneous program impacts and to control for selection at various stages (in dropout decisions, school-choice decisions, and grade retention).

6.4 *The unobserved type distributions*

Table 17 shows the distribution of the unobserved types, which is permitted to vary by *Prospera*-beneficiary status and by a measure of local poverty called the marginality index. Non-*Prospera* students are more likely to be type I or type II, whereas *Prospera* students are more likely to be type III or type IV. Table 18 explores how types are related to outcomes by simulating 9th-grade test scores and dropout outcomes when all *Prospera* beneficiaries are assumed to be of one type. Type I has the highest academic achievement on average (both in mathematics and Spanish), followed by types II and III. Type IV has the lowest academic achievement. Thus, we find the weaker types are disproportionately more likely to be enrolled in the *Prospera* program, indicating the presence of negative selection on unobserved factors in addition to the selection that is controlled by the observed characteristics.

Table 17. The unobserved type distributions.

	Type proportions			
	Type I	Type II	Type III	Type IV
Non-Prospera, High Marginality	0.12	0.47	0.24	0.17
Prospera, High Marginality	0.24	0.23	0.24	0.30
Non-Prospera, Low Marginality	0.45	0.19	0.21	0.15
Prospera, Low Marginality	0.20	0.14	0.49	0.18

If we simulate the average 9th-grade test-score impacts assuming that the type distribution for *Prospera* beneficiaries is the same as that of nonbeneficiaries (as indicated in the column labeled "Avg P=0"), we find that the positive effects of *Prospera* participation are underestimated for both mathematics (by 8.7 points) and Spanish (by 2.8 points). The decrease in dropout rates is also mitigated. Thus, selection on unobserved factors is an important feature of the data.

6.5 *Exploration of longer-term Prospera effects*

Because our data tracks students for a maximum of 6 years, we cannot observe longer-term outcomes for our baseline cohort (4th graders in 2008). We can, however, use data for a slightly older cohort—6th graders in 2007—to draw inferences about how program impacts on 9th-grade test scores relate to 12th-grade test scores and graduation rates, subject to some caveats described below. We estimate the following regression model:

$$y_{i,12} = \alpha_0^y + \alpha_1^y ENLACE_{i,9}^{\text{math}} + \alpha_2^y ENLACE_{i,9}^{\text{Spanish}} + \alpha_3^y \text{female} + \epsilon_{i,9}^y,$$

where $y_{i,12}$ consists of three different outcomes: (i) math test score at the end of grade 12; (ii) Spanish test score at the end of grade 12; and (iii) an indicator for whether completed secondary school (inferred from taking ENLACE test at end of grade 12). This regression is estimated using a sample of public-school students who took the 9th-grade ENLACE exams in 2010 and also have a 12th-grade ENLACE score.

Our previous findings showed that the *Prospera* program increases math scores by 0.21 standard deviations (SD) and Spanish scores by 0.04 SD by grade 9. We can obtain

Table 18. Test-score and dropout outcomes by unobserved types.

Outcomes	Type I	Type II	Type III	Type IV	Avg $P = 1$	Avg $P = 0$	Diff
Math score	601	582	553	522	561	570	−8.7
Spanish score	501	488	481	458	481	484	−2.8
Dropout	0.26	0.24	0.28	0.28	0.27	0.26	0.009

Note: Columns "Type I" to "Type IV" display the simulated test-score and dropout-rate outcomes if all *Prospera* students are assumed to be a particular type. The column "Avg P=1" reports the outcomes when *Prospera* students follow their own unobserved type distribution. The column "Avg $P = 0$" reports the alternative outcomes when *Prospera* students are assigned the unobserved type distribution for non-*Prospera* students. "Diff" reports the difference between "Avg $P = 0$" and "Avg $P = 1$."

Table 19. Longer-term extrapolation of *Prospera* effects to grade 12.

	Panel (a): full sample			Panel (b): twin subsample		
	(1) Math	(2) Spanish	(3) Grad.	(1) Math	(2) Spanish	(3) Grad.
+1 SD math (g9) (α_1^y)	0.34	0.12	0.068	0.23	0.16	0.034
	(0.002)	(0.002)	(0.0006)	(0.03)	(0.03)	(0.008)
+1 SD Spanish (g9) (α_2^y)	0.30	0.53	0.100	0.23	0.37	0.054
	(0.002)	(0.002)	(0.0006)	(0.03)	(0.03)	(0.008)
Twins FE	No	No	No	Yes	Yes	Yes
Prospera effect (Δ_{12}^y)	0.083	0.045	0.009	0.060	0.048	0.009

Note: Data Source: The ENLACE data set, as provided in De Hoyos, Estrada, and Vargas (2021). Our analysis focuses students in public schools who took the ENLACE exam in grade 9 in year 2010. Panel (a) uses all students while panel (b) focuses on twins (students who enrolled in the same school in grade 6, with identical last names and birth dates). Robust standard errors are reported in parentheses.

the predicted impact on grade 12 outcomes as follows:

$$\Delta_{12}^y = 0.21 * \alpha_1^{\text{math}} + 0.04 * \alpha_2^{\text{Spanish}}.$$

Panel (a) of Table 19 reports the estimated *Prospera* effects derived in this way. A 1-SD increase in 9th-grade math scores predicts a 0.34 SD increase in 12th-grade math scores and a 0.12 increase in 12th-grade Spanish scores. A 1-SD increase in the 9th-grade math scores is also associated with a 6.8 percentage point increase in the graduation rate. Similarly, a 1-SD increase in 9th-grade Spanish scores predicts a 0.30 SD increase in 12th-grade math scores, a 0.53 increase in 12th-grade Spanish scores, and a 10.0 percentage point increase in the graduation rate. Using the 9th-grade test score gains implied by our estimated *Prospera* impacts, we can infer that being a *Prospera* beneficiary from grades 4 through 9 would increase 12th-grade math scores by 0.083 SD, increase 12th-grade Spanish scores by 0.045 SD, and increase the upper-secondary-school graduation rate by 0.9 percentage points. These extrapolated longer-term *Prospera* impacts are likely underestimates, because they do not account for the fact that students typically get additional conditional cash transfers in grades 9–12, which are the grades with the largest transfers.[51]

A potential concern with using the above OLS regression model as a means of extrapolating longer-term impacts is the omission of family-background covariates, which could bias the estimated lagged test score coefficients. To address this concern, panel (b) performs a robustness check where we estimate the same model on a subsample of twins, adopting a twins-fixed-effects approach to control for interfamily differences in school, household, and neighborhood factors.[52] This estimation yields a slightly lower

[51] De Hoyos, Estrada, and Vargas (2021) demonstrate that higher 12th-grade ENLACE test scores predict better outcomes between ages 18–20. Specifically, they show that a one SD increase in these scores boosts the likelihood of university enrollment by 10% and, for employed individuals, leads to 6% higher wages within 2 years post-high school.

[52] Note we are not able to use the twins-sample approach to analyze the benefit of *Prospera*, because there is no variation in *Prospera* receipt for twins within the same household.

effect of the *Prospera* program on grade-12 math scores, but similar impacts on Spanish scores and graduation rates.

When interpreting our predicted *Prospera* effects at the end of grade 12, there are two important caveats. First, our estimated relationships between grades 9 and 12 test scores are based on individuals who participated in the ENLACE exams in both of those grades. This could lead to biased results due to selective attrition or selective exam-taking. Weaker students are more likely to drop out or repeat grades or otherwise not take the ENLACE exams and, therefore, are not accounted for in our estimations. Second, our extrapolation based on regression only assesses *Prospera*'s impact through its effect on 9th-grade test scores, potentially overlooking other program benefits, such as enhanced noncognitive skills or the advantages of continued exposure to *Prospera* during higher-secondary education. As a result, our predicted effects on outcomes in grade 12 probably represent lower bounds of the actual *Prospera* program impacts through grade 12.

7. Conclusions

Prior literature demonstrated substantial effects of the Mexican *PROGRESA/ Oportunidades/ Prospera* CCT program on educational attainment and schooling progression. However, little was known about how the program affected children's academic achievement, because comprehensive data to study that question were not available. Using newly available nationwide school-roster and test-score data, we develop and implement a model of school progression and academic achievements. Our modeling framework goes beyond the previous literature by integrating value-added test-score models, school-choice models that include the dropout option and account for local-labor-market-work opportunities, probabilistic grade retention, and measurement equations to allow for missing test-score data and to account for potential distortions arising from student cheating. Our analysis incorporates rich observed heterogeneity in family and child characteristics and unobserved heterogeneity in the form of discrete types, which enter multiple model equations. Our likelihood estimation approach explicitly controls for selective school enrollment/dropping-out and selection into different school types. Our estimation is based on multiple linked administrative and survey data sets as well as geocoded data on school locations, used to characterize individual-specific school-choice sets.

The data show that children from *Prospera*-beneficiary households live in less-urban areas and enroll in distance-learning schools (telesecondary) at much higher rates, so another goal of our analysis is to understand how academic achievement depends on the type of school attended. The question of whether a distance-learning modality is an effective way of teaching in comparison to fully in-person approaches is of considerable independent interest. Many countries face problems of how to provide access to high-quality schooling for students living in rural areas. There is scant evidence on the effectiveness of distance learning in such contexts.

Our key results include the following. First, the *Prospera* program did not substantially impact test scores in grades 5 and 6 in either general or indigenous schools. However, there are positive and statistically significant impacts on test scores in grades 7, 8, 9 with larger overall average impacts in mathematics (0.14–0.21 standard deviations) than in Spanish (0.04–0.06 standard deviations). The pattern of larger impacts at higher grade levels is perhaps to be expected given that the cash transfer amounts are significantly higher in lower-secondary than in primary grades. Second, we find that the *Prospera* program decreases the dropout rate by 7 percentage points, primarily at the 6th-to-7th-grade transition. Third, there is a consistent pattern of larger impacts for the most-disadvantaged children/youth, those in higher *Prospera*-participation propensity-score quartiles.

Fourth, the value-added parameter estimates indicate that lagged test scores are important determinants of current test scores, implying that *Prospera*-program effects on test scores accumulate over time. Our empirical findings show the importance of accounting for the dynamic nature of academic-achievement production to quantify the program gains accruing over multiple years. Fifth, we also find evidence of gender differences on test scores, with boys scoring higher on average on mathematics tests than girls and girls scoring higher on average on Spanish tests. The gender gaps in mathematics widen from primary to secondary grades. Participating in *Prospera* benefits leads to substantial test score increases for both girls and boys. The mathematics impacts are slightly greater for girls and the Spanish impacts are slightly greater for boys, which leads to a modest narrowing of the gender gaps in both subjects.

Sixth, our analysis takes into account that a small proportion of students in each grade were identified as having cheated (copied). Although cheating rates are low overall, they are somewhat higher in telesecondary schools and at higher grade levels. We develop an approach to account for possible test-score distortions (one-sided measurement error) arising from copying, which can affect both the dependent and right-side (lagged) variables in the value-added model.

Seventh, our analysis shows that telesecondary schools play an important role in the *Prospera*-program effect. Even with the copying adjustment, we find that telesecondary schools are at least similar in effectiveness and, in some cases, more effective than regular schools in teaching mathematics and Spanish. When we simulate the effects of removing the telesecondary school option for the children who attend these schools, their dropout rate prior to grade 9 increases from 18% to 52%, and educational attainment decreases by 1.2 grades. Thus, telesecondary schools are important to *Prospera*'s success in improving educational outcomes for disadvantaged students.

Lastly, we compared results based on our preferred model specification, integrating value-added and school-choice models and allowing for unobserved heterogeneity, to those obtained from a simpler model. Our extended approach led to estimates that in some respects, including program impacts, are substantially different. Overall, our results indicate that *Prospera* was not only effective in increasing school enrollments but also that the program led to significant positive impacts on academic achievement in mathematics and Spanish, with the most disadvantaged children/youth experiencing the greatest benefits and with distant learning through telesecondary schools playing critical roles.

References

Allende, Claudia, Francisco Gallego, Christopher Neilson et al. (2019), "Approximating the equilibrium effects of informed school choice." IRS Working Papers 628, Princeton University, http://arks.princeton.edu/ark:/88435/dsp01fj2364948. [0136]

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber (2005), "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools." *Journal of Political Economy*, 113 (1), 151–184. [0135, 0137]

Amemiya, Takeshi (1985), *Advanced Econometrics*. Harvard University Press. [0159]

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer (2002), "Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment." *American Economic Review*, 92 (5), 1535–1558. [0136]

Angrist, Joshua, Eric Bettinger, and Michael Kremer (2006), "Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia." *American Economic Review*, 96 (3), 847–862. [0136]

Arellano, Manuel and Stephen Bond (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *Review of Economic Studies*, 58 (2), 277–297. [0160]

Attanasio, Orazio P., Costas Meghir, and Ana Santiago (2012), "Education choices in Mexico: Using a structural model and a randomized experiment to evaluate progresa." *Review of Economic Studies*, 79 (1), 37–66. [0134, 0157]

Avitabile, Ciro and Rafael De Hoyos (2018), "The heterogeneous effect of information on student performance: Evidence from a randomized control trial in Mexico." *Jounal of Development Economics*, 135, 318–348. [0140]

Baird, Sarah, Francisco H. G. Ferreira, Berk Özler, and Michael Woolcock (2014), "Conditional, unconditional and everything in between: A systematic review of the effects of cash transfer programs on schooling outcomes." *Journal of Development Effectiveness*, 6 (1), 1–43. [0138]

Barham, Tania, Karen Macours, and John A. Maluccio (2013), "More schooling and more learning?: Effects of a three-year conditional cash transfer program in Nicaragua after 10 years." InterAmerican Development Bank. [0138]

Behrman, Jere R., Zednik Hrubec, Paul Taubman, and Terrence J. Wales (1980), *Socioeconomic Success: A Study of the Effects of Genetic Endowments, Family Environment, and Schooling*, Vol. 128. North-Holland Publishing Company, Amsterdam; New York. [0137]

Behrman, Jere R., Susan W. Parker, and Petra E. Todd (2005), "Long-term impacts of the oportunidades conditional cash transfer program on rural youth in Mexico." IAI discussion papers, University of Goettingen, https://www.econstor.eu/bitstream/10419/27412/1/504478389.PDF. [0134]

Behrman, Jere R., Susan W. Parker, and Petra E. Todd (2009), “Schooling impacts of conditional cash transfers on young children: Evidence from Mexico.” *Economic Development and Cultural Change*, 57 (3), 439–477. [0134]

Behrman, Jere R., Susan W. Parker, Petra E. Todd, and Weilong Zhang (2024), “Supplement to “prospering through Prospera: A dynamic model of cct impacts on educational attainment and achievement in Mexico”.” Quantitative Economics Supplemental Material. [0139]

Behrman, Jere R. and Mark R. Rosenzweig (1999), ““Ability” biases in schooling returns and twins: A test and new estimates.” *Economics of Education Review*, 18 (2), 159–167. [0137]

Behrman, Jere R., Piyali Sengupta, and Petra Todd (2005), “Progressing through PROGRESA: An impact assessment of a school subsidy experiment in rural Mexico.” *Economic Development and Cultural Change*, 54 (1), 237–275. [0134, 0138]

Bernal, Raquel and Michael P. Keane (2010), “Quasi-structural estimation of a model of childcare choices and child cognitive ability production.” *Journal of Econometrics*, 156 (1), 164–189. [0162]

Boardman, Anthony E. and Richard J. Murnane (1979), “Using panel data to improve estimates of the determinants of educational achievement.” *Sociology of Education*, 52 (2), 113–121. [0135]

Borghesan, Emilio and Gabrielle Vasey (2024), “The marginal returns to distance education: Evidence from Mexico’s telesecundarias.” *American Economic Journal: Applied Economics*. [0139]

Bravo, David, Sankar Mukhopadhyay, and Petra E. Todd (2010), “Effects of school reform on education and labor market performance: Evidence from Chile’s universal voucher system.” *Quantitative Economics*, 1 (1), 47–95. [0136]

Cameron, Stephen V. and James J. Heckman (1998), “Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males.” *Journal of Political Economy*, 106 (2), 262–333. [0135, 0137]

Cameron, Stephen V. and James J. Heckman (2001), “The dynamics of educational attainment for black, Hispanic, and white males.” *Journal of Political Economy*, 109 (3), 455–499. [0135, 0172]

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff (2014a), “Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates.” *American Economic Review*, 104 (9), 2593–2632. [0135]

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff (2014b), “Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood.” *American Economic Review*, 104 (9), 2633–2679. [0135]

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov (2006), “Interpreting the evidence on life cycle skill formation.” *Handbook of the Economics of Education*, 1, 697–812. [0135, 0136, 0137]

Cunha, Flavio and James J. Heckman (2008), "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation." *Journal of Human Resources*, 43 (4), 738–782. [0153]

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach (2010), "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica*, 78 (3), 883–931. [0135, 0136]

De Hoyos Navarro, Rafael E., Orazio Attanasio, and Costas Meghir (2019), "Can scholarships increase high school graduation rates? Evidence from a randomized control trial in Mexico." Policy Research Working Paper 8826, World Bank, https://hdl.handle.net/10986/31581. [0140]

De Hoyos, Rafael, Ricardo Estrada, and María José Vargas (2018), "Predicting individual wellbeing through test scores: Evidence from a national assessment in Mexico." The World Bank. [0140]

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, 105524. [0175]

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. [0140]

Epple, Dennis, Akshaya Jha, and Holger Sieg (2018), "The superintendent's dilemma: Managing school district capacity as parents vote with their feet." *Quantitative Economics*, 9 (1), 483–520. [0136]

Fabregas, Raissa and Laia Navarro-Sola (2023), "Broadcasting education at scale: The long-term effects of Mexico's telesecundarias." Working paper, URL, https://cega.berkeley.edu/wp-content/uploads/2020/03/Fabregas_PacDev2020.pdf. [0139]

Figlio, David N. and Cecilia Elena Rouse (2006), "Do accountability and voucher threats improve low-performing schools?" *Journal of Public Economics*, 90 (1-2), 239–255. [0136]

Fiszbein, Ariel and Norbert R. Schady (2009), *Conditional Cash Transfers: Reducing Present and Future Poverty*. World Bank Publications. [0134, 0138]

Fox, Jeremy T., Kyoo il Kim, Stephen P. Ryan, and Patrick Bajari (2012), "The random coefficients logit model is identified." *Journal of Econometrics*, 166 (2), 204–212. [0159]

Gallego, Francisco and Andrés Hernando (2009), "On the determinants and implications of school choice: Structural estimates and simulations for Chile." Documento de Trabajo, 343. [0135, 0136]

Glewwe, Paul (2002), "Schools and skills in developing countries: Education policies and socioeconomic outcomes." *Journal of Economic Literature*, 40 (2), 436–482. [0135]

Hadna, Agus Heruanto and Dyah Kartika (2017), "Evaluation of poverty alleviation policy: Can conditional cash transfers improve the academic performance of poor students in Indonesia?" *Cogent Social Sciences*, 3 (1), 1295548. [0138]

Hanushek, Eric (1979), "Conceptual and empirical issues in the estimation of educational production functions." *Journal of Human Resources*, 14, 351–388. [0135]

Hastings, Justine, Thomas J. Kane, and Douglas O. Staiger (2009), "Heterogeneous preferences and the efficacy of public school choice." Vol. 2145, 1–46. NBER working paper. [0136]

Hastings, Justine S., Christopher A. Neilson, and Seth D. Zimmerman (2012), "The effect of school choice on intrinsic motivation and academic outcomes." Working Paper 18324, National Bureau of Economic Research, URL http://www.nber.org/papers/w18324. [0136]

Heckman, James J., John Eric Humphries, and Gregory Veramendi (2016), "Dynamic treatment effects." *Journal of Econometrics*, 191 (2), 276–292. [0156, 0158, 0159]

Heckman, James J., John Eric Humphries, and Gregory Veramendi (2018), "Returns to education: The causal effects of education on earnings, health, and smoking." *Journal of Political Economy*, 126 (S1), S197–S246. [0157, 0159]

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd (1997), "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme." *Review of Economic Studies*, 64 (4), 605–654. [0153, 0154]

Heckman, James J. and Salvador Navarro (2007), "Dynamic discrete choice and dynamic treatment effects." *Journal of Econometrics*, 136 (2), 341–396. [0157, 0159, 0160]

Heckman, James J. and Burton Singer (1984), "A method for minimizing the impact of distributional assumptions in econometric models for duration data." *Econometrica*, 52, 271–320. [0137, 0153]

Heckman, James J. and Edward Vytlacil (2001), "Policy-relevant treatment effects." *American Economic Review*, 91 (2), 107–111. [0156]

Heckman, James J. and Edward Vytlacil (2005), "Structural equations, treatment effects, and econometric policy evaluation 1." *Econometrica*, 73 (3), 669–738. [0156]

Heckman, James J. and Edward J. Vytlacil (2007), "Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation." *Handbook of Econometrics*, 6, 4779–4874. [0156]

Honoré, Bo E. and Ekaterini Kyriazidou (2000), "Panel data discrete choice models with lagged dependent variables." *Econometrica*, 68 (4), 839–874. [0159]

Hsieh, Chang-Tai and Miguel Urquiola (2006), "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program." *Journal of Public Economics*, 90 (8-9), 1477–1503. [0136]

Kane, Thomas J., Daniel F. McCaffrey, Trey Miller, and Douglas O. Staiger (2013), "Have we identified effective teachers? Validating measures of effective teaching using random assignment." Research Paper. MET Project. Bill & Melinda Gates Foundation, Citeseer. [0135]

Kane, Thomas J. and Douglas O. Staiger (2008), "Estimating teacher impacts on student achievement: An experimental evaluation." Working Paper 14607, National Bureau of Economic Research, URL http://www.nber.org/papers/w14607. [0135]

Keane, Michael P., Petra E. Todd, and Kenneth I. Wolpin (2011), "The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications." In *Handbook of Labor Economics*, Vol. 4, 331–461, Elsevier. [0157]

Leite, Phillippe G., Ambar Narayan, and Emmanuel Skoufias (2011), "How do ex ante simulations compare with ex post evaluations? Evidence from the impact of conditional cash transfer programs." Policy Research working paper WPS 5705, World Bank, https://hdl.handle.net/10986/3469. [0157]

Manski, Charles F. (1988), "Identification of binary response models." *Journal of the American statistical Association*, 83 (403), 729–738. [0159]

McEwan, Patrick J. (2001), "The effectiveness of public, catholic, and non-religious private schools in Chile's voucher system." *Education Economics*, 9 (2), 103–128. [0135]

Mroz, Thomas A. (1999), "Discrete factor approximations in simultaneous equation models: Estimating the impact of a dummy endogenous variable on a continuous outcome." *Journal of Econometrics*, 92 (2), 233–274. [0153]

Navarro-Sola, Laia (2019), "Secondary school expansion through televised lessons: The labor market returns of the Mexican telesecundaria." HCEO working paper, University of Chicago, https://humcap.uchicago.edu/RePEc/hka/wpaper/NavarroSola_2021_secondary-schools-televised-lessons.pdf. [0171]

Neal, Derek (1997), "The effects of catholic secondary schooling on educational achievement." *Journal of Labor Economics*, 15 (1, Part 1), 98–123. [0135]

Parker, Susan W. and Petra E. Todd (2017), "Conditional cash transfers: The case of progresa/oportunidades." *Journal of Economic Literature*, 55 (3), 866–915. [0134, 0143, 0152, 0157]

Parker, Susan W. and Tom Vogl (2023), "Do conditional cash transfers improve economic outcomes in the next generation? Evidence from Mexico." *The Economic Journal*, 133 (655), 2775–2806. https://doi.org/10.1093/ej/uead049. [0134, 0138]

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain (2005), "Teachers, schools, and academic achievement." *Econometrica*, 73 (2), 417–458. [0137]

Rothstein, Jesse (2009), "Student sorting and bias in value-added estimation: Selection on observables and unobservables." *Education Finance and Policy*, 4 (4), 537–571. [0137]

Rouse, Cecilia Elena (1998), "Private school vouchers and student achievement: An evaluation of the Milwaukee parental choice program." *Quarterly Journal of Economics*, 113 (2), 553–602. [0136]

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. [0135]

Schellenberg, Jonathan and Christopher R. Walters (2020), "Do parents value school effectiveness?" *American Economic Review*, 110 (5), 1502–1539. [0136]

Schultz, T. Paul (2004), "School subsidies for the poor: Evaluating the Mexican Progresa poverty program." *Journal of Development Economics*, 74 (1), 199–250. [0134, 0138]

SEP (2010), "Enlace: Manual tecnico 2010." Working paper, URL http://red.sevalladolid.mx/pdf/20150721121435997529manual_tecnico_enlace10.pdf. [0140, 0146]

SEP (2018a), "Cuestionarios de contexto (enlace)." https://xaber.org.mx/repositorio-de-datos/, last accessed on 2018-11-15. [0140]

SEP (2018b), "Resultados de aprendizaje (enlace)." https://xaber.org.mx/repositorio-de-datos/, last accessed on 2018-11-15. [0141]

Skoufias, Emmanuel and Susan W. Parker (2001), "Conditional cash transfers and their impact on child work and schooling: Evidence from the progresa program in Mexico." *Economia*, 2 (1), 45–96. [0137]

Snilstveit, Birte, Emma Gallagher, Daniel Phillips, Martina Vojtkova, John Eyers, Dafni Skaldiou, Jennifer Stevenson, Ami Bhavsar, and Philip Davies (2017), "Protocol: Interventions for improving learning outcomes and access to education in low-and middle-income countries: A systematic review." *Campbell Systematic Reviews*, 13 (1), 1–82. [0138]

Summers, Anita A. and Barbara L. Wolfe (1977), "Do schools make a difference?" *American Economic Review*, 67 (4), 639–652. [0135]

Todd, Petra E. and Kenneth I. Wolpin (2003), "On the specification and estimation of the production function for cognitive achievement." *Economic Journal*, 113 (485), F3–F33. [0135, 0137]

Todd, Petra E. and Kenneth I. Wolpin (2006), "Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility." *American Economic Review*, 96 (5), 1384–1417. [0134, 0138, 0157]

Co-editor Limor Golan handled this manuscript.

Manuscript received 22 November, 2022; final version accepted 19 September, 2024; available online 9 October, 2024.

The replication package for this paper is available at https://doi.org/10.5281/zenodo.13771826. The Journal checked the data and codes included in the package for their ability to reproduce the results in the paper and approved online appendices.