IDB WORKING PAPER SERIES Nº 833

The Consequences of Educational Voucher Reform in Chile

Richard J. Murnane
Marcus R. Waldman
John B. Willett
Maria Soledad Bos
Emiliana Vegas

Inter-American Development Bank
Education Division

August 2017

The Consequences of Educational Voucher Reform in Chile

Richard J. Murnane*
Marcus R. Waldman*
John B. Willett*
Maria Soledad Bos**
Emiliana Vegas**

* Harvard University Graduate School of Education
** IDB Education Division

August 2017

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

The consequences of educational voucher reform in Chile
/ Richard J. Murnane, Marcus R. Waldman, John B. Willett, María Soledad Bos, Emiliana Vegas.
p. cm. — (IDB Working Paper Series ; 833)
Includes bibliographic references.

1. Educational vouchers-Chile-Econometric models. 2. School choice-Chile-Econometric models. 3. Education and state-Chile-Econometric models. 4. Academic achievement-Chile-Econometric models. I. Murnane, Richard J. II. Waldman, Marcus R. III. Willett, John B. IV. Bos, María Soledad. V. Vegas, Emiliana.

VI. Inter-American Development Bank. Education Division. VII. Series.
IDB-WP- 833

http://www.iadb.org

Abstract[*]

In an effort to boost student achievement and reduce income-based gaps, the Chilean government passed the *Preferential School Subsidy Law* (SEP) in 2008, which altered the nation's 27-year-old universal school-voucher system dramatically. Implementation of SEP increased the value of the school voucher by 50 percent for "priority students", primarily those whose family incomes fell within the bottom 40 percent of the national distribution. To be eligible to accept the higher-valued vouchers from these students, schools were required to waive fees for Priority students and to participate in an accountability system.

Using national data on the mathematics achievement of 1,631,841 Chilean 4th grade students who attended one of 8,588 schools during the years 2005 through 2012, we address two research questions (RQs):

1. Did student test scores increase and income-based score gaps become smaller during the five years after the passage of SEP?
2. Did SEP contribute to increases in student test scores and, if so, through what mechanisms?

We addressed these RQs by fitting a sequence of multi-level interrupted time-series regression models, supplemented by other descriptive analyses. We found that:

1. On average, student test scores increased markedly and income-based gaps in those scores declined by one-third in the five years after the passage of SEP.
2. The combination of increased support of schools and accountability was the critical mechanism through which the implementation of SEP increased student scores, especially in schools serving high concentrations of low-income students. Migration of low-income students from public schools to private voucher schools played a small role.

We conclude by responding to a recent paper by Feigenberg, Rivkin, and Yan (2017) that argues that the gains from SEP are illusory.

[*] The authors would like to acknowledge the helpful comments on earlier drafts provided by Cristian Bellei, Gregory Elacqua, Christopher Jencks, and Norbert Schady.

I. Introduction

Debates about the merits of market-based strategies to improve student achievement have a long history in both the USA and internationally. In 1962, University of Chicago economist Milton Friedman argued that a universal voucher system would improve both the quality and the efficiency of the U.S. K-12 education system. Under his proposal, parents of school-aged children would receive a voucher that they could use to pay part, or all, of the cost of enrolling their child in a private school. Competition among schools for students would improve the quality of American education.

Writing almost a decade later, Christopher Jencks (1970) argued that vouchers do indeed have the potential to improve educational outcomes, especially for economically disadvantaged children, but only if the system has a very different design than that which Friedman described. Jencks proposed a system in which vouchers provided to low-income families would have greater value than those given to higher-income families, and the admission and dismissal procedures of participating private schools would be highly regulated.

Over subsequent decades, economists developed a number of theoretical models that describe how universal vouchers would influence both the distribution of students among schools and the distribution of student achievement. These models highlight the potential importance of the density of nearby educational options, the role of peer groups, the value of the vouchers for families with particular characteristics, and rules regarding the admission and dismissal procedures of participating schools.[1] To date, however, there have been no opportunities in the United States to examine the importance of these design elements in large-scale universal voucher systems empirically.[2] For such evidence, we must turn to Chile.

I. Educational Vouchers in Chile

In 1981, Chile introduced a universal educational voucher system for students in both its elementary and secondary schools. At the same time, the central government transferred the administration of public schools to municipal governments. Since economists from the University of Chicago advised the Chilean government, it is not surprising that Chile's voucher system bore similarities to the design that Friedman had proposed.

---

1 See, for example, Epple & Romano (1998) and Nechyba (2000, 2006).

2 A number of studies have examined the impacts on student achievement of targeted voucher programs in the United States. See Epple, Romano, & Urquiola (2017) for a review of the evidence up to 2014, and Abdulkadiroğlu, Pathak, & Walters (forthcoming), and Dynarski et al. (2017) for newer evidence.

Key elements of the voucher system included:

a. Three types of schools served children, including public schools funded by voucher receipts, private schools financed by voucher receipts (henceforth, private voucher schools), and private schools that did not participate in the voucher system and that were financed by fees parents paid. Both for-profit and not-for profit organizations operated private schools.

b. The financial value of the voucher did not depend on family income.

c. Private voucher schools could decide which students to admit. Public schools were obligated to accept all students.

d. Public schools and private schools had substantial flexibility in hiring teachers and deciding how much to pay them.

e. A national system of standardized assessment of students' academic skills (SIMCE) was implemented to provide parents with comparative information about the achievement of students enrolled in different schools.

The basic design of the voucher system in Chile remained in effect through 2007, with two notable exceptions. After the restoration of democracy to the country in 1990, the salaries of public school teachers were increased and uniform salary schedules, which based pay on seniority and credentials, were restored. These changes, reflected in a new "Teacher Statute," affected only public-school teachers. In 1993, the Chilean government responded to fiscal pressures by introducing a system of "shared financing," under which private voucher schools were permitted to charge all parents fees in addition to the value of the voucher. The percentage of private voucher schools that charged fees rose rapidly, and more than half did so in 2007. The average fee for schools serving elementary-school students in that year was $30 per month, with a maximum of $121 per month. The value of the voucher was discounted for schools charging fees greater than one-half the value of the voucher.[3]

The introduction of the voucher system elicited a number of responses. The percentage of students enrolled in public schools declined markedly, from 78 percent in 1980 to less than

[3] Elacqua et al. (2016) and Bellei & Vanni (2015) provide descriptions of the voucher system and of changes in Chilean educational policies over the last three decades. Elacqua (2012, p.450, footnote 18) explains that the value of the voucher was discounted by 10 percent for schools that charged fees that were one-half to one times the value of the voucher. The discount rate was 20 percent for schools that charged fees greater than the value of the voucher.

50 percent in 2007. The percentage of students, especially those from middle-class families, enrolled in private voucher schools grew substantially. Many low-income parents also enrolled their children in private voucher schools. However, this did not result in an increase in school integration by socioeconomic status because private voucher schools tended to specialize. Some charged substantial fees and enrolled students from middle-class families.[4] Others charged either low or no fees and served students from low-income families primarily (Contreras, Sepulveda, and Bustos, 2010). The net effect was that school segregation by socioeconomic status increased substantially in the first two decades of the voucher system. Moreover, student achievement in mathematics and Spanish language, as measured on national tests, did not increase (Hsieh and Urquiola, 2006; Elacqua, 2012; Valenzuela, Bellei, & de los Rios, 2013; Epple, Romano, & Urquiola, 2017).

At the turn of the 21st century, student achievement in Chile was low relative to that of students in other countries participating in international test-score comparisons (Gonzales et al., 2000), and family-income based gaps in student achievement were large. These patterns contributed to the impetus for the substantial educational reforms that the Chilean government enacted in 2008.

Changes in the Voucher System

With the primary goals of decreasing inequality in student achievement and segregation among schools by socioeconomic status, the Chilean national government passed the Preferential School Subsidy Law (SEP) in January 2008. This landmark legislation made the Chilean educational voucher system more like the regulated compensatory voucher model that Christopher Jencks had proposed. SEP recognized explicitly that it costs more to educate students from low-income families well, especially in schools serving large percentages of children living in poverty. Under SEP, the vouchers provided to “priority students,” basically, those whose families were in the bottom 40 percent of the income distribution, were worth 50 percent more than those provided to other students. In addition to the higher-valued vouchers, schools serving large percentages of Priority students received per-student concentration bonuses, the size of which increased as the percentage of Priority students in the school’s student body increased.

[4] Children from affluent families were likely to attend high-tuition private schools that did not participate in the voucher system.

To be eligible to receive the higher-valued vouchers and concentration bonuses, schools had to agree to participate in the SEP program. One program requirement was that schools could not charge fees to priority students, although private voucher schools could do so for non-priority students. A second requirement was that participating schools had to agree not to select students based on their academic skills, nor expel them on academic grounds.

A third requirement was that schools had to participate in an accountability system that, for the first time, made schools responsible for the use of financial resources and student test scores. The Chilean Education Ministry classified schools participating in SEP as Autonomous, Emerging, or Recovering, depending on their students' scores on the national assessment and other performance indicators. Schools in the lower two categories had less autonomy in allocating their SEP resources than did autonomous schools. Schools in lower-ranked categories received support from the Education Ministry in drafting their progress plans and technical assistance in carrying them out. Struggling schools that failed to improve their students' mathematics and reading scores after receiving assistance risked losing their license or their eligibility for the higher-valued vouchers provided to Priority students.

When SEP was launched in 2008, it covered preschool through 4th grade, and one additional grade was added to the coverage in each subsequent year.[5] Almost all public schools and about two-thirds of private subsidized elementary schools chose to participate in SEP in 2008.[6] Those that did were free to use the extra resources they received for serving priority students to improve the education of all students. Consequently, SEP may have benefitted non-priority students.

II. Research Questions

As Epple, Romano, & Urquiola (2017) explain, it is difficult to produce unbiased estimates of the causal impacts of changes in a national program. Several recent studies, which to our knowledge are unpublished, used different strategies in attempting to do so. These studies informed our work. Carrasco (2014) used a comparative interrupted time-series approach to estimate the impact of SEP on the mathematics and Spanish language achievement of 4th-grade

[5] It is noteworthy that the introduction and later expansion of SEP occurred during two different political administrations, with differing political views.

[6] In 2011, the government modified SEP in several ways: (a) extending benefits to middle-school students; (b) increasing the value of vouchers Priority students received; and (c) allowing schools greater flexibility in using government funding. In previous years, schools could not spend more than 15 percent of the SEP resources on personnel and had restrictions on the number of extra-hours they could pay their teachers. These constraints were removed in 2011.

students in Chile. He did so by comparing “the deviation from prior outcome trends among a ‘treatment group’ that received the extra SEP funds to the analogous deviation from a ‘comparison group’ that did not receive these extra resources” (p.9). He found that four years of SEP participation increased 4th-grade students’ mathematics achievement by 0.18 standard deviations “compared to students in schools that did not participate in the policy” (p. 10). One critical assumption underlying the validity of Carrasco’s approach is that the deviation from prior outcome trends for the comparison group provides an unbiased estimate of the counterfactual–that is, what the deviation in outcome trends would have been for schools participating in SEP had they chosen not to do so. Since schools weighed the benefits and costs of deciding whether to participate in SEP, this assumption may not be valid.

Neilson (2015) examined how SEP influenced the distribution of student achievement within the context of a demand-and-supply model of school choice. In Neilson’s model, spatially differentiated schools compete for students by offering combinations of quality and price. Families make schooling choices by comparing the quality/price combinations offered by schools in their neighborhood. Neilson used detailed data on school fees and locations to fit his hypothesized statistical model. He then used the obtained parameter estimates to simulate how the changes in school prices that SEP provided to low-income families affected schooling choices and the distribution of student achievement. He found that SEP increased the test scores of low-income students by 0.20 standard deviations and closed the income-based achievement gap by one-third.

As is often the case with highly structured approaches to policy analysis, Neilson made several decisions in developing and fitting his statistical model that may have influenced his results and their interpretation. One was to characterize SEP in terms of a policy that changed the schooling prices that low-income families faced. This depiction allowed Neilson to incorporate the impact of SEP into his supply-and-demand model. However, it meant downplaying the accountability requirements that schools participating in SEP were subject to, and that may have had a marked impact on the performance of their students. A second decision was to assume that peer effects were not important. This assumption reduced the complexity of Neilson’s supply-and-demand model substantially. However, recent studies using compelling research designs (e.g., Carrell & Hoekstra, 2010; Carrell, Hoekstra, & Kuka, 2016) demonstrate that peer groups have substantial and lasting impacts on classmates’ academic success.[7]

[7] Neilson (2015) noted that he is developing a version of his model that will include peer group influences.

Navarro-Palau (2016) used two sources of variation in schooling options to analyze the impact of SEP on the enrollment choices and mathematics and Spanish language achievement of groups of 4th-grade students, defined by their mother's education level. The first is the timing of the introduction of SEP. The second is exogenous variation in the timing of school entry stemming from the age cutoff in Chile for entry into first grade. Using a regression-discontinuity and difference-in-differences framework, she found that the greatest impact of SEP on school choice occurred among Priority students with relatively well-educated mothers. Passage of SEP increased the percentage of children with mothers who had completed high school that enrolled in private voucher schools that did not charge fees, but did not increase the percentage of Priority students with less-educated parents who did so. Navarro-Palau found that the impact of SEP on student achievement was modestly positive, with the greatest gains going to Priority students enrolled in public schools.

One of the strengths of Navarro-Palau's paper is the distinction she makes between private voucher schools that charge fees and those that do not. We go a step further and distinguish between for-profit and not-for-profit private voucher schools that charge fees and those that do not. This distinction matters because legislation passed in 2015 mandates that only not-for-profit organizations are eligible to operate private schools that receive vouchers. A second difference between our paper and Navarro-Palau's is that we explicitly incorporate characteristics of the nested structure of the data into our statistical modeling, with students clustered within schools and years, and schools being observed for as many as 8 years. As we explain below, this model structure allows us to test quite detailed hypotheses about differences between the pre-SEP and post-SEP periods in the distribution of school performance trends.

In our research, we addressed two research questions:

RQ1: Trends in Student Test scores? Did student scores increase and income-based score gaps become smaller during the five years after the passage of SEP?

RQ2: Role of SEP? Did SEP contribute to increases in student test scores and, if so, through what mechanisms?

III. Research Design

Dataset

We analyzed administrative data that the Chilean government collected annually on school characteristics and student enrollments, family characteristics as reported on parental surveys, and the results of nationally normed and year-to-year equated standardized tests that assessed the mathematics achievement of all students in Grade 4. We focused exclusively on this grade because it was the highest one included in SEP's initial year of implementation. We merged these datasets, matching on student and school IDs. The resulting dataset contained information on every student enrolled in Grade 4 in a Chilean public or private school, in each year from 2005 through 2012.

A typical Chilean school contributed three years of student test-score data before the initiation of the SEP program (2005-2007), and five years of data thereafter (2008-2012).[8] Thus, our data are longitudinal at the school level, implying that testing instances (henceforth referred to as "testing year") are nested within schools. However, our data are not longitudinal at the student level because we use only information on the test scores of students who were in Grade 4, in each school, in each year. Thus, students are nested within a school and within a testing year, a nesting reflected explicitly in the error-covariance structure that we have specified in all our subsequent statistical models.[9]

Sample

In constructing our analytic sample, we excluded from the dataset: (a) the twelve percent of students for whom 4th-grade test scores were not available, (b) the less than five percent of students who were enrolled in special schools for children with disabilities, (c) the less than one percent of students enrolled in schools for which organizational type was not available, and (d) the seven percent of children enrolled in high-fee private schools that did not participate in the voucher system.

[8] The school year in Chile runs from March through December. The Chilean legislature passed the SEP legislation in January 2008 and consequently SEP was in operation during the 2008 school year.

[9] Two and one-half percent of students repeated 4th grade and therefore appear more than once in our analytic sample. We included an indicator coded to identify these students as a covariate in all our statistical models. In effect, we treat "grade repeaters" as separate students in the same school at different years. Moreover, 0.05 percent of students have two test scores recorded in the same year at the same school. We eliminated from our analytic sample all but one of the records for each of these students.

In Table 1, we provide selected summary statistics on our sample for 2005, the baseline year. We list the number and percentage of elementary schools of each organizational type that enrolled 4th-grade students in that year (excluding elite private schools that did not participate in the voucher system). We distinguish among five school types. The first are public schools. We have classified the remaining (private) schools that accepted government vouchers into four groups, defined by their characteristics in 2007: (a) for-profit schools that did not charge fees, (b) for-profit schools that did charge fees, (c) non-profit schools that did not charge fees, and (d) non-profit schools that did charge fees. We list the number and percentage of 4th-grade students enrolled in each type of Chilean elementary school, along with summary statistics on selected characteristics of these students. More than three-fifths of the 6,871 elementary schools that accepted educational vouchers in 2005 were public schools. Twenty-six percent of the schools were for-profit private organizations, and slightly more than half of these charged fees in addition to the value of the voucher. Eleven percent of the schools were not-for-profit private organizations, and slightly more than half of these charged fees.

Characteristic	All Schools	School Organizational Form				
		Public Schools	Private Schools			
			No-Fee		Fee-Charging	
			Not For-Profit	For-Profit	Not For-Profit	For-Profit
Number and % of Schools Serving Grade-4 Students	6,871 (100%)	4,311 (62.7%)	345 (5.0%)	789 (11.5%)	411 (6.0%)	1,015 (14.8%)
Number and % of Grade-4 Students	220,521 (100%)	118,377 (53.7%)	13,295 (6.0%)	13,830 (6.3%)	27,811 (12.6%)	47,208 (21.4%)
Average Family-Income Percentile	47.4	39.1	45.9	38.2	64.9	60.8
Median Father's Educational Attainment	12	9-11	9-11	9-11	12	12
Median Mother's Educational Attainment	12	9-11	9-11	9-11	12	12
Average Student Mathematics Score	244.6	235.1	247.3	227.1	270.0	257.9

Table 1. Baseline (in 2005) Characteristics of Chilean Schools and Their 4th-Grade Students, Nation-Wide and by School Type.

Not surprisingly, in 2005, the distribution of 4th-grade students across school types is similar in many respects to the distribution of the school types. Slightly more than half of Chilean 4th-grade students attended public schools in 2005. Another 21 percent attended for-profit private schools that charged fees, and 13 percent attended not-for-profit private schools that charged fees. Approximately 12 percent of 4th-grade students attended private schools that did not charge fees, with about half of these enrolled in for-profit private schools and the other half in not-for-profit schools.

Students enrolled in public schools or private schools that did not charge fees came from families that were considerably less advantaged than students enrolled in private schools that charged fees. On average, their parents had lower educational attainments and lower family incomes, reflecting the significant sorting of students by socioeconomic status that Hsieh & Urquiola (2006) noted. Given this sorting, it is not surprising that children enrolled in public schools or no-fee private schools had lower average mathematics scores than did children enrolled in fee-charging private schools.

In Table 2, we provide selected descriptive statistics on our sample separately for schools and students in rural and urban areas. The striking differences between the characteristics of rural and urban students and the schools they attended led us to hypothesize that the impacts of SEP might be different for rural students than for those living in cities. The reason is that, due to low population density, rural families have many fewer schooling options than urban families do. Consequently, we incorporated the urban/rural distinction explicitly in our statistical models.

Notice that there were more public schools in rural areas (2,444) than in urban areas (1,867) in 2005, even though there were five times as many 4th-grade students living in urban areas (98,424) than the number living in rural areas (19,953). One factor contributing to this pattern is that 82 percent of rural students attended public schools while only 50 percent of urban students did so. Another is that the rural public schools had much lower enrollments (an average of eight grade-4 students per school) than urban public schools (an average of 53 grade -4 students per school).

Rural students were also distributed differently across private-school types than urban students were. For instance, 38 percent of urban 4th-grade students attended fee-charging private schools, while less than two percent of rural students did so. In fact, there were only seven fee-charging for-profit private schools in rural areas in 2005 and the same number of fee-charging not-for profit private schools. In contrast, private for-profit schools that did not charge fees served a larger percentage of rural 4th-grade students (11) than urban 4th-grade students (6).

Characteristic	Urban Settings					Rural Settings				
	Public Schools	Private Schools				Public	Private Schools			
		No-Fee		Fee-Charging			No-Fee		Fee-Charging	
		Not For-Profit	For-Profit	Not For-Profit	For-Profit		Not For-Profit	For-Profit	Not For-Profit	For-Profit
Number and % of Schools Serving Grade-4 Students	1,867 (48.5%)	228 (5.9%)	345 (9.0%)	404 (10.5%)	1,008 (26.2)	2,444 (81.0%)	117 (3.9%)	444 (14.7%)	7 (0.2)	7 (0.2)
Number and % of Grade-4 students	98,424 (50.1%)	12,087 (6.2%)	11,121 (5.7%)	27,560 (14.0%)	47,092 (24.0%)	19,953 (82.3%)	1,209 (5.0%)	2,708 (11.2%)	249 (1.0%)	112 (0.5%)
Average Number of Grade-4 Students Per School	52.7	53.0	32.2	68.2	46.7	8.2	10.3	6.1	35.6	16.0
Average Family-Income Percentile	41.2	47.4	41.8	65.1	60.8	28.7	30.5	23.7	51.3	68.0
Median Mother's Educational Attainment	9-11	12	9-11	12	12	8	8	0-7	12	>=13
Average Student Mathematics Score	236.0	249.7	231.9	270.2	257.9	230.9	223.8	207.7	249.9	266.1

Table 2. Baseline (in 2005) Characteristics of Chilean Elementary Schools and Their 4th Grade Students, By Location and School Type.

Note: Excluded students include those attending high-fee private schools that did not participate in the voucher system, those attending schools for students with disabilities, those attending schools for which the organizational form was not specified, and those for whom test were not available.

Another important difference between rural and urban 4th-grade students is that rural students lived in more economically disadvantaged families. On average, their parents had lower educational attainments and less income. For these reasons, it is not surprising that rural 4th-grade students scored lower, on average, on the national mathematics examination (228) than did urban grade-4 students (247).

Measures

To keep the presentation of our findings brief, we report and discuss only the results of fitting models in which the outcome is either a student's observed score on the national mathematics test (*MATH*, RQ1) or the same score adjusted for the influence of selected family-background characteristics (*ADJ_MATH*, RQ2). We focus on mathematics achievement, instead of Spanish-language achievement, because U.S.-based studies have found that the mathematics achievement of young children is a stronger predictor of later academic success (Duncan & Magnuson, 2011) and of subsequent labor-market outcomes (Murnane, Willett, & Levy, 1995) than their language achievement. However, in additional analyses in which we replicated our analyses with language achievement as the outcome, our results were qualitatively similar, although the average differences in achievement among students in different school types and locations were somewhat smaller. Over the period from 2005-2012, scores on the national standardized mathematics test for Chilean 4th-grade students enrolled in public schools or in private voucher schools ranged from 74 to 395, with a mean of 249 and a standard deviation of approximately 53 points.

We included in all models (a) an integer variable representing the chronological year (*YEAR*), and (b) a dichotomous variable indicating whether the SEP program was operating in that year (*SEP*, coded 1 for years 2008 through 2012; 0 otherwise). With one exception that we explain below, we centered predictor *YEAR* on 2008, the first year in which the SEP program operated.

Other predictors of student mathematics achievement include selected characteristics of students and their families, and schools. Forty-nine percent of the students are female, and are designated by the dichotomous predictor, *FEMALE* (0=male; 1=female). Among students living in urban areas, the average achievement of male students (252) was approximately 4 points (0.08 S.D.) higher than that of female students during the pre-SEP years 2005-2007. Among students living in rural areas, the gender gap had the same direction, but was smaller in magnitude, with the average score of females (233) one point lower than that of males. Slightly less than four percent of the students in the sample repeated Grade 4. They are differentiated by the dichotomous indicator *REPEATER* (0= not repeated; 1=repeated).

We also included in our statistical models selected family characteristics as predictors of 4th-grade students' test scores. They included: (a) the educational attainments of the mother and father, and (b) family income. We treated the lowest educational category ("Some Elementary School") as the omitted (reference) category for both mother's and father's educational attainment. We then included, as predictors, dichotomous indicators to distinguish the four higher levels of attainment of mothers and fathers: (a) Elementary School Graduate; (b) Some High School; (c) High-School Diploma; and (d) At Least Some Post-Secondary Education, each coded 1 to indicate the presence of the category concerned, 0 otherwise. Over the entire period of observation, the median educational attainment of both fathers and mothers was a high-school graduate.

Parents were asked to report, via a survey, into which of a number of pre-designated ranges their family income fell.[10] Using the method described in Reardon (2011), we estimated the percentile of the family-income distribution of Chilean 4th graders into which each family's income fell in that year, resulting in the continuous covariate *INC*. The reason that we chose to control for family-income percentile, rather than family income is that the former provides a common metric across the years of observation. In our analyses, we found that the relationship between student mathematics achievement and family-income percentile was described parsimoniously by a third-order polynomial specification. As our family-income and parental-educational attainment predictors contained missing values, we used multiple imputation to deal with the non-responses, (Rubin, 1987).[11]

We also included as covariates in our analyses two sets of school characteristics. One consists of a vector of dichotomous variables that describe the type of school, distinguishing the four types of private voucher schools described above. In our statistical models, we treated public school as the omitted (reference) category. The second was a dichotomous variable, *RURAL*, that distinguished schools located in rural areas from those located in rural areas. In 2005, 56 percent of Chilean elementary schools (public and private) that accepted educational vouchers were located in urban areas of the country (*RURAL*=0). In our initial statistical models, we included interaction terms that permitted different parameter values for the impact on achievement of student and family characteristics and school characteristics for rural and for urban students. In our final models, we retained only those interactions with location that proved statistically significant.

[10] We created our continuous measure of family-income percentile by converting responses to a measure of family income in the parent survey that required parents to respond in one of 13 ordinal categories in the years from 2005 to 2008, and 15 categories in the years 2008 to 2011. The categories do not provide detailed information on the low end of the family income distribution, especially in the early years.

[11] In our dataset, between 16 percent and 33 percent of the values of mother's education were missing, depending on the year. Correspondingly, between 20 percent and 35 percent of the values of father's education were missing, again depending on year. Additionally, in most years, between 16 percent and 23 percent of the values of family income were missing, with no discernable trend in the rate across years, and in 2007, 32 percent of the values of this variable were missing. We used the method of multiple imputation to eliminate the missing values and mitigate bias, fitting our hypothesized models in eight multiply imputed data sets and pooling parameter estimates across datasets using Rubin's rules.

Data-Analyses

Our primary analytical strategy was to fit a sequence of multi-level interrupted time-series models built around the same core specification, in which either MATH or ADJ_MATH (depending on the research question) was hypothesized to be a function of the main effects of YEAR and SEP, and their interaction. Each model included other predictors, the choice of which depended on the research question. As noted earlier, all our statistical models incorporated an error-covariance structure for the random effects that reflected the complex hierarchical structure of the data, with students nested in schools and schools contributing student test data for as many as eight years. To simplify exposition, in the text, we present only the composite models resulting from this specification rather than documenting specifically the full complexity of the hypothesized error structure. We present the corresponding complete specification of the hypothesized multi-level models in Appendix A. We fit all of the interrupted time-series models using the MIXED routine in Version 14 of Stata.

A strength of our modeling strategy is that it allowed us to use general linear hypothesis testing to test subtle hypotheses about differences in trends in the distribution of student achievement during the pre-SEP and post-SEP periods. For example, we were able to test whether the variation in school performance trends after the passage of SEP differed from that during the pre-SEP years.

One limitation of our interrupted time-series models is that they do not provide a convincing test of whether SEP *caused* the increase in average student test scores and the closing of the income-based average test-score gap that we describe below. An alternative explanation is that other influences on students and/or schools contemporaneous with the implementation of SEP caused the changes in the test-score distribution. To eliminate this alternative explanation, we would have needed data on a comparable group of students that experienced these “other influences,” but were not subject to SEP implementation. We believe that no such legitimate comparison group exists. However, with our design and data, we are able to examine whether test scores rose more during the post-SEP years for students attending the types of schools most influenced by SEP. We also present evidence on the likely consequences for student test scores of other changes in the lives of Chilean students and the schools they attended.

RQ1: Trends in Student Achievement? To address our first research question, we specified the following interrupted time-series model, for the k^{th} 4th-grade student enrolled in the i^{th} school in the j^{th} year.

$$
\begin{aligned}
MATH_{ijk} = {} & \gamma_{00} + \gamma_{10}(YEAR_j - 2008) \\
& +\gamma_{20}SEP_j + \gamma_{30}[SEP_j \times (YEAR_j - 2008)] \\
& +\boldsymbol{\beta}_1'\boldsymbol{INC}_{ijk} + \boldsymbol{\beta}_2'[\boldsymbol{INC}_{ijk} \times (YEAR_j - 2008)] \\
& +\boldsymbol{\beta}_3'(\boldsymbol{INC}_{ijk} \times SEP_j) \\
& + \boldsymbol{\beta}_4'[(\boldsymbol{INC}_{ijk} \times SEP_j) \\
& \times (YEAR_j - 2008)] + \epsilon_{ijk}'
\end{aligned}
\qquad [1]
$$

where ϵ_{ijk}' represents a complex hypothesized error term that embodies the nested hierarchical structure of the analytic sample. (See Appendix A for a full specification of the error-covariance structure).

The model includes time-varying predictors to describe the main effect of YEAR and SEP, and their two-way interaction. This specification permits unique population average trends in achievement over time in the pre- and post-SEP periods. Then, to permit these trajectories to differ by family income, we have also included the main effect of vector $\boldsymbol{INC}$ (containing the student's family-income percentile, its square, and its cube). Our model also contains cross-product terms representing the two-way interactions between **INC** and YEAR, and between **INC** and *SEP*, plus the three-way interaction among all three predictors. This specification permits the hypothesized relationship between student mathematics achievement and both time and family income percentile to differ in the pre- and post-SEP periods.

After fitting this hypothesized "full" model, we relied on judicious simultaneous hypothesis testing to remove unneeded terms from the model. We then interpreted parameter estimates from the final reduced model, and used them to reconstruct and display average trends in mathematics achievement over time and by family-income percentile, in both the pre- and post-SEP periods, for prototypical students.

RQ2: Role of SEP?

The mechanisms through which the implementation of SEP could alter the distribution of student mathematics scores include facilitating the movement of students to schools with superior instruction and/or academically stronger peer groups and by improving the instruction in schools through a combination of increased financial resources and greater accountability for student test scores. We used several strategies to assess the relative importance of these mechanisms.

First, we examined whether groups of schools defined by organizational type and location that had the highest participation rate in SEP also had the greatest increase in students' average mathematics scores during the post-SEP period. In addressing this question, we considered it critical to control for the effects of selected important student and family characteristics (student gender, whether a student had repeated 4th grade, family income, and parental educational attainment). This is because it may have been easier for schools to improve their average student achievement post-SEP by attracting more advantaged students than by improving the quality of the education they provided. However, we recognized that an explicit goal of SEP was to reduce achievement gaps based on socioeconomic status. If SEP succeeded, the parameters associated with these student and family covariates would have smaller values in the post-SEP than in the pre-SEP period. As explained in Appendix A, we adopted a two-step procedure to adjust each student's mathematics score for the influences of student and family background influences, while attributing to SEP reductions in the influences of these variables on student achievement. All subsequent analyses treated the adjusted mathematics score, *ADJ_MATH,* as the outcome.

We fit the following multilevel interrupted time-series model, for the k^{th} 4th-grade student enrolled in the i^{th} school in the j^{th} year:

$$\begin{aligned} ADJ\_MATH_{ijk} = {} & \gamma_{00} + \boldsymbol{\gamma_o'} \boldsymbol{W_i} + \gamma_{10}(YEAR_j - 2008) \\ & + \boldsymbol{\gamma_1'}[\boldsymbol{W_i} \times (YEAR_j - 2008)] + \gamma_{20} SEP_j + \boldsymbol{\gamma_2'}[\boldsymbol{W_i} \times SEP_j] \\ & + \gamma_{30}\{SEP_j \times (YEAR_j - 2008)\} \\ & + \boldsymbol{\gamma_3'}[\boldsymbol{W_i} \times SEP_j \times (YEAR_j - 2008)] + \boldsymbol{\epsilon''_{ijk}} \end{aligned} \qquad [2]$$

where ϵ''_{ijk} is a composite multilevel time-dependent residual.

In Model 2, we include the time-invariant predictor vector $\boldsymbol{W}$ and its interactions with time and SEP status in order to distinguish among schools based on their organizational type and location. After fitting this "full" model, we again used judicious tests of simultaneous statistical inference to prune unnecessary terms, leading to a more parsimonious final model.

We supplement our first strategy for assessing the role of SEP and the importance of specific mechanisms (the fitting of our hypothesized interrupted time-series models) with other descriptive analyses. Second, we examined whether implementation of SEP altered the long-term decline in the percentage of Chilean elementary school students enrolled in public schools and the long-term increase in the percentage enrolled in for-profit private voucher schools. One reason this might have occurred is that passage of SEP resulted in a dramatic increase in funding for public schools, most of which served large percentages of children from low-income families. On the other hand, the higher-valued vouchers that SEP provided to low-income parents may have increased their access to private schools, and could have accelerated the trend away from enrollment in public schools.

Third, we used the method described by Oaxaca (1973) to decompose the increase between 2006 (a pre-SEP year) and 2012 (the fifth year after passage of SEP) in the average adjusted achievement of 4th-grade students in the bottom half of the family-income distribution (henceforth, low-income students) into three components.[12] The logic underlying this descriptive decomposition is that it sheds light on the relative importance of improvement in performance of schools of each type and changes in the distribution of students among school types in accounting for the increase in the average adjusted achievement of low-income 4th-grade students between 2006 and 2012.

Fourth, we measured the extent to which the pattern of segregation of low-income 4th-grade students into different schools from those attended by higher-income students was different in 2012 than it was in 2006. Here, we used a method described by Clotfelter (2004) to decompose the segregation of low-income students in each year into several parts, each the result of a hypothetical experiment.

In describing our results, we refer frequently to *rates of school improvement*. We use this term to mean changes over time in the average adjusted mathematics achievement of 4th-grade

[12] Our preferred strategy in analyzing the mechanisms through which SEP altered the distribution of adjusted student achievement was to compare the distribution of this outcome for students in 2007, the last pre-SEP year, and in 2012, the last post-SEP year for which we have data. However, in some cases, we wanted to compare results for low-income students in a pre-SEP and post-SEP year. This required a common definition of low-income students. In each year, parents were asked to report in which of a number of pre-specified ranges their family income fell. The distribution of responses across bins in 2007 and 2012 made it impossible to define “low-income” in terms of a common income percentile. In contrast, it was possible to define “low-income” using a quite comparable metric in 2006 and 2012.

students attending particular schools. Our valid use of this term rests on two assumptions. The first is that the background characteristics of individual students that we have included in our statistical models control adequately for any direct effect of year-to-year changes in the student body on the average achievement of grade-4 students in that school. The second is that improvements in the mathematics scores of students in a particular school do not come at the cost of foregone improvements in other dimensions of students' skills and knowledge. Finally, we want to emphasize that schools, as we use the term, are complex organizations in which adults with varying capabilities and incentives work together to enhance children's skills and knowledge. Unfortunately, we lack the data to shed light on the ways that implementation of SEP altered the characteristics of school staffs, their capabilities and incentives, and the ways that they interact with children.

IV. Results

RQ1: Trends in Student Test Scores

In Figure 1, we display the predicted average mathematics score of 4th-grade students by family-income percentile at the end of the 2005 school year, two years before the passage of the SEP legislation, and at the end of the 2012 school year, five school years after the passage of the legislation. We derived the plotted values in the figure from the estimates of the parameters of Equation [1], which are listed in Appendix B, Table B1.

Notice three patterns in Figure 1. First, the fitted curves sloping upward from left to right show the strong role of family income in predicting the mathematics score for 4th-grade students in Chile. For example, in 2005, the predicted score of students whose families were at the 85th percentile of the income distribution (249) was 21 points (0.4 SD) higher than the predicted score for students from families at the 15th percentile of the income distribution.[13]

The second pattern, illustrated by the vertical distance between the fitted 2005 and the 2012 curves, is the increase between 2005 and 2012 in the predicted mathematics score for students at every family income percentile. For example, the increase in the predicted mathematics score for students at the 15th percentile of the family-income distribution was 16 points (approximately 0.32 SD).

[13] We did not report the conventional 90-10 income percentile gap because the income bins in the questionnaire that parents completed did not distinguish among incomes at the top of the distribution.

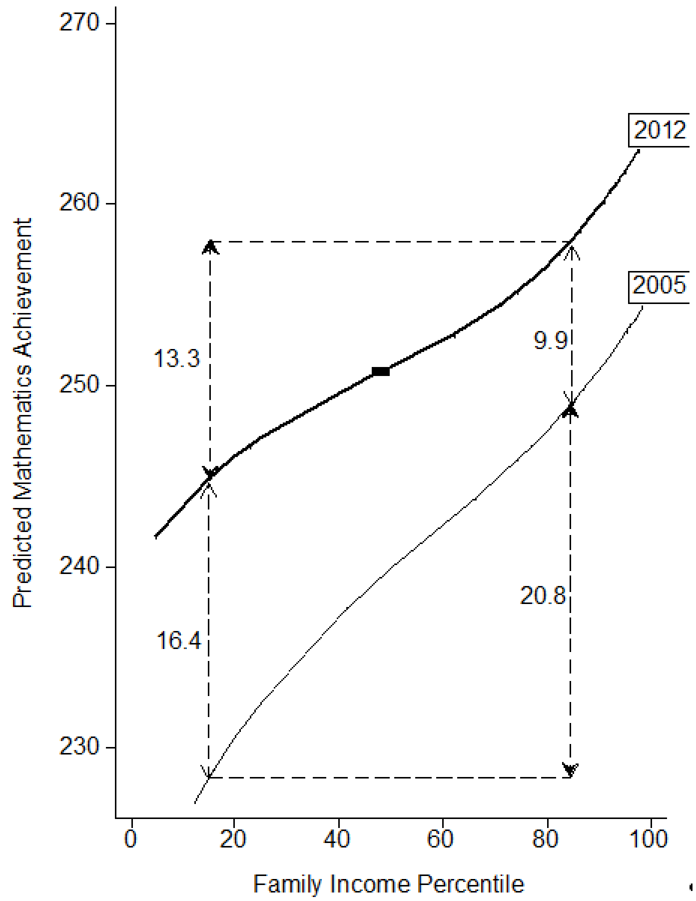

Figure 1: Predicted mathematics scores for prototypical 4th-grade students by family income percentile in 2005 and in 2012.

The third pattern is the decline between 2005 and 2012 in the size of income-based gaps in mathematics test scores (p<0.01). This is the consequence of the larger increase in mathematics achievement for students at the bottom of the family-income distribution than for those at the top. For instance, the gap between the predicted mathematics achievement of students from the 15th and 85th family-income percentiles declined from 21 points in 2005 to 13 points in 2012.

In summary, between 2005 and 2012 the mathematics scores of Chilean 4th-grade students increased substantially, and the size of the income-based test-score gap in mathematics declined by at least one-third.

RQ2: Role of SEP

Was the post-SEP rate of performance improvement greater for schools located in cities, where the density of private schools was substantial, than for schools located in rural areas, where schools typically faced relatively little competition for students? Did post-SEP performance improve the most in the types of schools that had served primarily low-income students during the pre-SEP years? We addressed these questions by fitting the multilevel model specified in Equation 2. We present estimated parameters from this fitted model in Appendix B, Table B2, along with corresponding standard errors, approximate *p*-values and goodness-of-fit statistics. Figures 2, 3, and 4 are based on the parameters of this fitted model.

School Location

In Figure 2, we display fitted trends in average adjusted mathematics scores for prototypical 4th-grade students who attended public schools in either rural areas (dashed line) or urban areas (solid line). (Approximately 50 percent of grade-4 students living in urban areas and 82 percent of those living in rural areas attended public schools in 2005.) Note that the average adjusted mathematics score for 4th-grade students who attended rural public schools lies below that of similar students who attended urban public schools throughout the period of observation from 2005 through 2012. One potential explanation for the consistent but modest difference in average adjusted achievement is that it is difficult to attract skilled teachers to schools located in rural areas.

A second pattern illustrated in Figure 2 is that the average adjusted mathematics scores of 4th-grade students in both urban and rural public schools declined over the pre-SEP years, 2005-2007, at a rate of about two points (0.04 SD) per year. As in any discontinuity design, the extensions of these pre-SEP fitted lines to year 2008 provide predictions of what the average adjusted mathematics scores would have been in 2008 for 4th-grade students enrolled in urban and rural public schools had SEP not been introduced before the start of the 2008 school year.

A further striking pattern shown in Figure 2 is that the average adjusted mathematics scores of 4th-grade students enrolled in both urban and rural public schools rose markedly during the five years following the passage of SEP. The first-year impacts are shown in Figure 2 by the vertical distance on the plot separating the right endpoint of the pre-SEP adjusted-score projection and the corresponding left end-point of the post-SEP adjusted-score trend-line. These initial impacts, 3.1 and 1.7 points for students in the rural and urban schools respectively, are small. This is not surprising as schools that chose to participate in SEP in 2008 had almost no time to plan how to use the additional school funding that came along with the enrollment of low-income students, and then to implement those plans effectively.

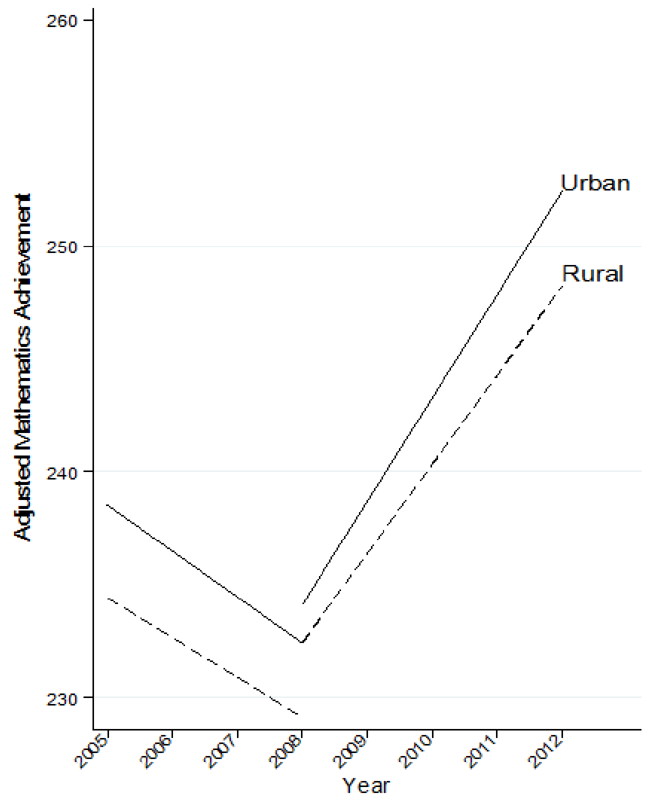

Figure 2: Fitted trends in average adjusted mathematics scores for prototypical 4th-grade students who attended public schools in rural or urban areas.

During the four-year period from the end of 2008 to the end of 2012, the average adjusted mathematics score of 4th-grade students enrolled in urban public schools increased at an annual rate of 4.6 points annually (0.09 SD) and that of students enrolled in rural public schools increased at a rate of 4.0 points annually (0.08 SD). Note that we specified the post-SEP trajectories as linear in time in our statistical model (Equation 2). We found no evidence that the respective rates of increase in students' adjusted mathematics scores lessened over this period.[14]

There are several potential explanations for the steady improvement in students' adjusted math scores during the first five years after the passage of the SEP legislation. First, schools may have needed several years to learn how to use additional resources productively. Second, the

---

14 We fit additional statistical models that included the quadratic effects of year in the post-SEP performance trends. We found no evidence of a negative coefficient on any of the additional quadratic terms.

amount of additional funds schools received for serving low-income students increased over the period of observation. Third, the percentage of the nation’s students qualifying for higher-valued vouchers rose from approximately 40 percent in 2008 to more than 50 percent in 2012. Fourth, it took the Chilean government several years to implement fully the accountability provisions that schools participating in SEP must abide by. Finally, in 2011, the government passed legislation that made all schools receiving government funding, even those that chose not to participate in SEP, accountable for demonstrating improvement in student test scores. While we cannot assess the relative importance of these factors in contributing to the improvement in student achievement during the post-SEP period, together their influences were substantial. At the end of the 2012 school year, five years after the passage of SEP, the average adjusted mathematics score of students attending public schools, either in urban or rural areas, was more than one-third of a standard deviation higher than that of their counterparts attending public schools five years earlier.

School Type

In examining trends in school performance for schools with different organizational forms, we focus on those located in urban areas. The reason is that there were almost no fee-charging private schools in rural areas. Moreover, we could not reject the null hypothesis that the average rates of performance improvement for the two types of private schools located in rural areas were the same as that of rural public schools. In contrast, there was more variation in the organizational forms of elementary schools and in their performance trajectories.

We display in Figure 3 fitted trends in performance for the five types of urban schools that accepted government vouchers. Performance declined during the pre-SEP years in four of the five groups of schools (the exception being private no-fee, for-profit schools, in which adjusted achievement in the baseline year was particularly low). Average adjusted mathematics achievement increased during the post-SEP period for all five types of schools.

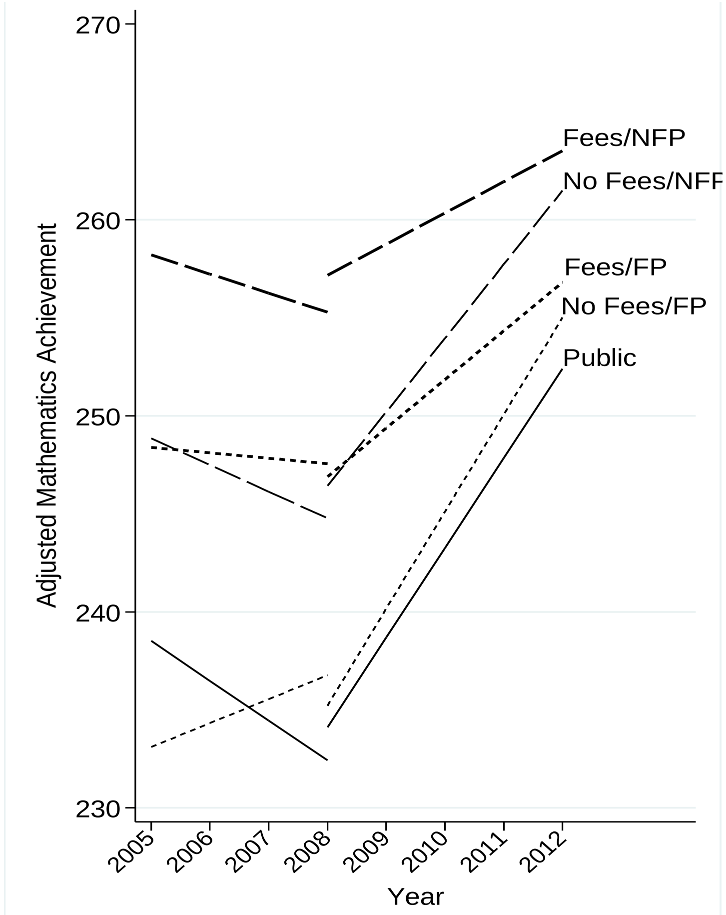

Figure 3: Fitted trends in average adjusted mathematics scores for prototypical 4th-grade students who attended one of five types of schools in urban areas.

A closer inspection of Figure 3 reveals more subtle patterns. First, focus on the pre-SEP period. Average adjusted mathematics scores were highest for 4th-grade students who were enrolled in not-for-profit private schools that charged fees. Next highest were the performance profiles of students who attended either for-profit private schools that charged fees or not-for-profit private schools that did not charge fees. Finally, students enrolled in either public schools or for-profit schools that did not charge fees had the lowest performance profiles. Differences in resource levels are a likely explanation for these differences in average adjusted scores. Private schools that charged fees in addition to receiving the value of government-provided vouchers had higher per-student revenues than did the public schools and the for-profit private schools that did not charge fees. The not-for-profit private schools that did not charge fees may have garnered additional revenue from charitable contributions.

A second pattern illustrated in Figure 3 is that average performance estimates for the five types of urban elementary schools were closer to each other in 2012, five years after the passage of SEP, than they were in the pre-SEP years. This is a direct consequence of heterogeneity in the average rates at which schools of different types improved their performances during the post-SEP period. In particular, the average annual rates of performance improvement were greater in schools that did not charge fees (4.6 points in public schools; 3.8 points in private no-fee not-for-profit schools; 5.0 points in private no-fee for-profit schools) than in the two types of private school that had charged fees in 2007 (1.6 points annually in fee-charging not-for-profit schools; 2.5 points annually in fee-charging for-profit schools) (p<0.01).

One explanation for this pattern is that private schools that charged fees were less likely to choose to participate in SEP than were schools that did not charge fees. As we illustrate in Figure 4, almost all public schools chose to participate in SEP in 2008, and consequently benefitted from its provisions. Among private voucher schools, more than two-thirds of those that had *not* charged fees in 2007 joined SEP in 2008, and by 2012 approximately 90 percent of these schools were participating. In contrast, only about 40 percent of private schools that *had* charged fees in 2007 chose to participate in SEP in 2008, and by 2012, only about half had done so. This pattern suggests that the combination of additional funding and greater accountability brought on by the implementation of SEP was a key mechanism through which the SEP program improved 4th-grade student outcomes. This pattern is consistent with recent evidence from the USA showing that the combination of increased funding and accountability resulted in improved student test scores (Jackson, Johnson, & Persico, 2016).

The striking patterns in pre-SEP and post-SEP performance trends displayed in Figures 2 and 3 raise questions about patterns of variability in performance trends among schools. For instance, one might ask: Did heterogeneity in performance trends among schools of the same type differ between the pre-SEP and post-SEP periods? We can manipulate the estimated random-effects parameters (variances and covariances) displayed at the bottom of Appendix B, Table B2 algebraically to address such questions. For instance, the estimated between-school variance of school-specific rates of improvement in average adjusted mathematics scores is 10.35 in the post-SEP period, within school type. This is 50 percent larger than the corresponding between-school variance of 7.15 in the pre-SEP period. Thus, not only did average trends in adjusted mathematics performance differ markedly between the pre- and post-SEP periods, there was also greater between-school variation in school-specific performance trends in the post-SEP period than pre-SEP, within each school type.

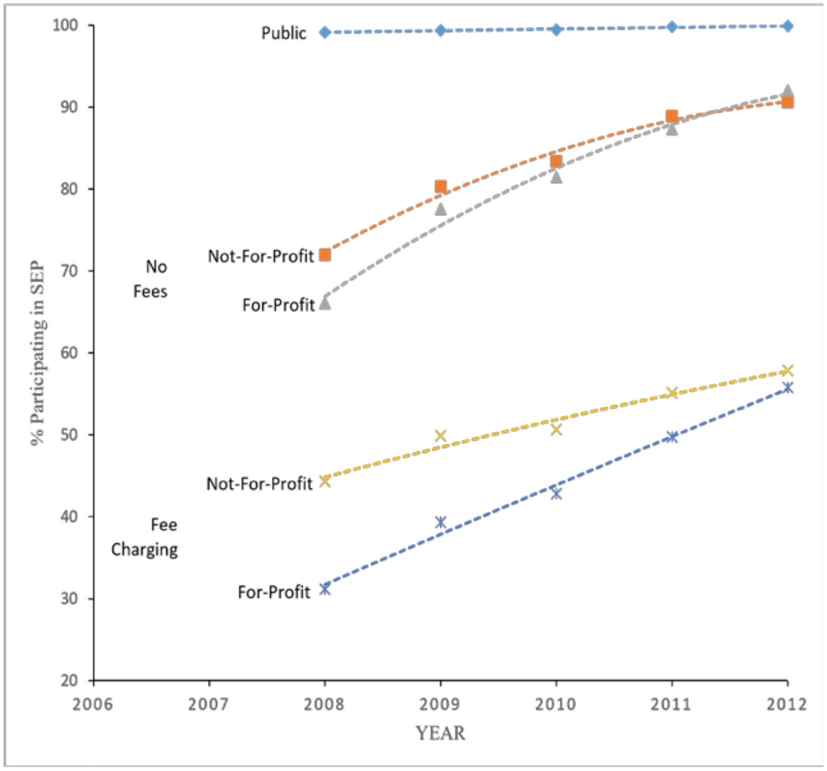

Figure 4: Fitted trends in the percentage of Chilean elementary schools, by type, that chose to participate in SEP.

Correspondingly, we can manipulate the random-effects estimates in Appendix B, Table B2 to reveal interesting patterns in the between-school correlations between the school-specific pre- and post-SEP rates of improvement in performance and their performance in 2005, the baseline year. For instance, the estimated correlation of the pre-SEP school-specific rates of performance improvement and the performance level in 2005 is positive (0.55), within school-type. This means that gaps in performance between the best- and worst-performing schools of each type widened during the pre-SEP period.

In contrast, the correlation between the post-SEP rate of improvement in school-specific performance and the performance level in 2005 is negative (-0.35). This indicates that gaps in performance between the best- and worst-performing schools of each type narrowed during the first five years after the passage of SEP.

Several factors probably contributed to the larger variance in school-specific rates of performance improvement in the post-SEP period than in the pre-SEP period, within school-type. First, since schools differed in the percentage of low-income students they served, the amount of extra funding they received as a result of joining SEP also differed. Second, some schools may have been much more effective in using their extra funds to improve student performance and to respond to accountability pressures than were other schools.

One plausible explanation for the negative covariance in the post-SEP period between school-specific rates of performance change and performance level in the baseline year is the influence of the accountability system to which schools participating in SEP became subject. This system targeted schools in which student test scores were especially low, and therefore these schools faced the greatest pressure to improve their students' performance. Later in the paper we consider whether the test scores gains reflect increases in students' skills and knowledge.

Trends in School-Enrollment Rates

As illustrated in Figure 2, the average adjusted mathematics achievement of 4th-grade students attending public schools in either urban or rural areas increased rapidly during the post-SEP period. One might predict that this substantial improvement in performance would have slowed the long-term migration of students away from public schools and toward private schools. However, this did not occur, as we illustrate in the two panels of Figure 5. In the left panel of the figure, we display trends in the number of 4th-grade students enrolled in each of the five types of schools that accepted vouchers. Notice that, between 2005 and 2012, the enrollment of 4^{th}-grade students in public schools declined steadily. Conversely, enrollment in the private schools that accepted vouchers either remained quite stable or increased, depending on the type of private school. The rescaling used in the right-hand panel of the figure makes these patterns easier to see. Here, we display the number of 4th-grade students attending each of the five different types of schools as a *percentage* of the number of students enrolled in each school type in 2005, the base year. Notice the dramatic increases in the number of students enrolled in private voucher schools that charge fees.

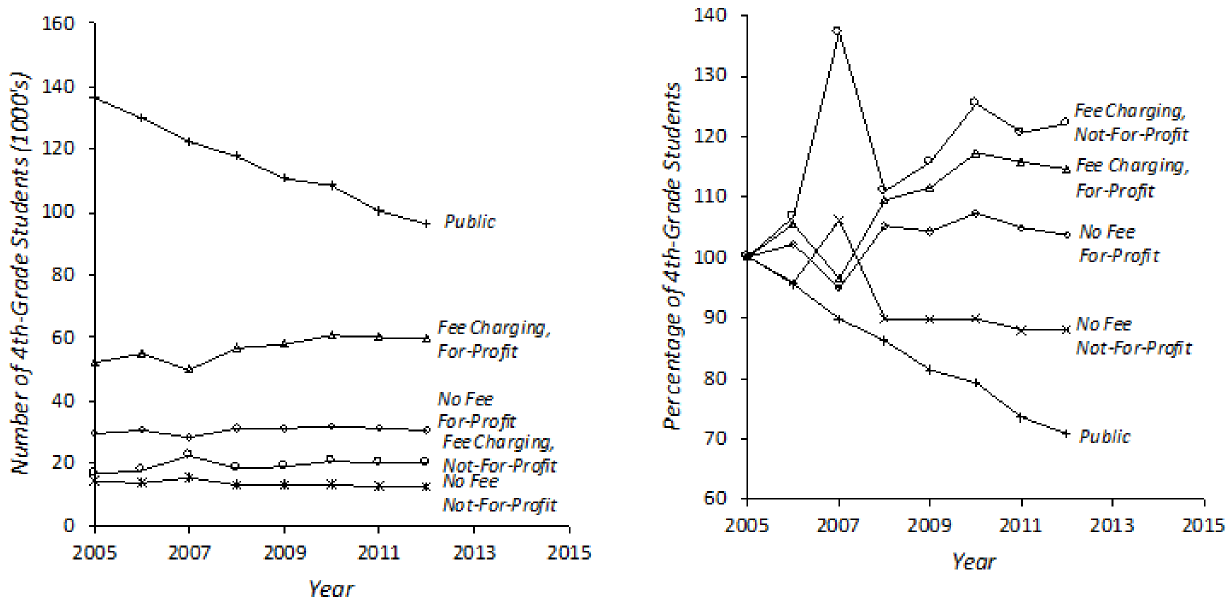

Figure 5: Trends over time, for 2005 through 2012, in the number (1000's, left panel) and proportion (right panel, expressed as a percentage of the number of students in baseline year 2005) of 4th-grade students enrolled in Chilean schools that participated in the voucher program.

One reason the improvement in the performance of public schools did not slow the migration of students from these schools may have been parents' responses to the school academic rankings that the Chilean government has published each year since 1995. These rankings are based on student test scores, unadjusted for the influences of student and family background characteristics. Since public schools in urban areas serve students from lower-income families, on average, than do private schools, the rankings of public schools remained low during the post-SEP period despite the substantial increases in the mathematics achievement of their students.

Decomposing Adjusted-Achievement Gains for Low-Income Students

The enrollment patterns displayed in Figure 5 raise the question of whether the shift of students away from public schools and into different types of private school accounted for a substantial part of the increase over the post-SEP period in the adjusted mathematics scores of Chilean 4th-grade students. Given that a goal of SEP was to reduce income-based gaps in achievement, we are especially interested in the factors contributing to the increase in the average adjusted mathematics achievement of low-income students. We used a method proposed by Oaxaca (1973) to decompose the increase in the average adjusted mathematics score of low-income students between 2006 and 2012 into three parts.[15] The first is a weighted average of the increases in the average adjusted achievement of students enrolled in each of the five types of schools. The second is a weighted average of the changes in average achievement stemming from differences between 2006 and 2012 in the distribution of students among the five types of schools. The third is a weighted average of interaction terms consisting of the products of changes in the average adjusted achievement and changes in the distribution of students. In our case, this third part was very small.

We found that more than 90 percent of the improvement in the average adjusted mathematics score of grade-4 low-income students between 2006 and 2012 stemmed from increases in the average achievement within each sector. These increases ranged from 12 points for private voucher schools that charged fees to 27 points for private no-fee for-profit schools. Changes in the distribution of low-income students across school types between 2006 and 2012 were substantial. They included a 12-point decline in the percentage enrolled in public schools and a seven-point increase in the percentage enrolled in private for-profit schools that charged fees. However, the changes in the distribution of students across sectors accounted for only seven to nine percent of the increase in the average adjusted mathematics score of low-income students between 2006 and 2012.[16]

Changes in the School Segregation of Low-Income Students

One mechanism through which SEP could have reduced income-based gaps in student achievement is by increasing low-income students' access to schools with higher-income, academically strong peer groups. To examine the extent to which this took place, we adapted an approach described by Charles Clotfelter (2004) to compare patterns of school segregation by income for 4th-grade students in 2006 and in 2012. Our measure of segregation is the difference between the overall proportion of low-income 4th-grade students in the country (n_k) and the proportion enrolled in the average higher-income 4th-grade student's school (E). In 2006, n_k = 0.56 and E= 0.39, so the value of the segregation measure is 0.17. In 2012, n_k = 0.59 and E = 0.42, so the value of the segregation measure for that year is also 0.17. Thus, by this measure, there was no change between 2006 and 2012 in the extent to which low-income 4th-grade

[15] We chose 2006 as the "initial" year for our Oaxaca decomposition rather than 2007, the last pre-SEP year, because it was not possible to adopt a common definition of "low-income student" for the years 2007 and 2012.

[16] In conducting an Oaxaca decomposition, it is necessary to decide which year to treat as the base year that is used in calculating the weights. The 7 percent figure comes from treating 2006 as the base year. The 9 percent figure comes from treating 2012 as the base year.

students were segregated into different schools from those attended by 4th-grade students from higher-income families.[17]

What did differ between the two years is the relative importance of the two factors that contribute to school segregation nationally: school segregation among 4th-grade low-income students living in the same commune, and segregation stemming from low-income 4th-grade students living in different communes from their higher income peers. We discovered this pattern by decomposing our measure of desegregation in each year into six constituent components through a set of successive steps, each representing a hypothetical redistribution of students. The results of this set of hypothetical exercises are displayed in Figure 6. In the figure, the light gray bars illustrate the percentage of school segregation by income that would be eliminated if low-income students were equally distributed among schools in a particular group in 2006. The dark gray bars provide the same information for 2012.

In the first step, all 4th-grade students enrolled in public schools in the same commune are redistributed so that each public school in the commune has the same share of 4th-grade low-income students. As illustrated by the top set of bars in Figure 7, this would reduce school segregation by income in 2006 by 20 percent, but only by 7 percent in 2012. This means that, on average, public schools in each commune were less socioeconomically segregated in 2012 than they were in 2006.

n the second step, all 4th-grade students in each commune that were enrolled in private voucher schools that had not charged fees in 2007 are redistributed such that each of these schools has the same percentage of 4th-grade low-income students. As illustrated by the second set of bars in Figure 6, this hypothetical step would reduce segregation by 6 percent in 2006 and by 3 percent in 2012. This means that no-fee private schools within each commune were less segregated by income in 2012 than in 2006.

In the next step, all 4th-grade students in each commune who were enrolled in private voucher schools that had charged fees in 2007 are redistributed to equalize the percentage of 4th-grade low-income students in each of these schools within each commune. As illustrated by the third set of bars in Figure 6, this step would reduce segregation by 25 percent in 2012, but only by 18 percent in 2006.

[17] Clotfelter's index of segregation, $S = (n_k - E)/n_k$, is slightly different from ours. We did not adopt Clotfelter's measure because the limitations of our measure of family income prevent us from determining precisely how the value of n_k differed between 2006 and 2012.

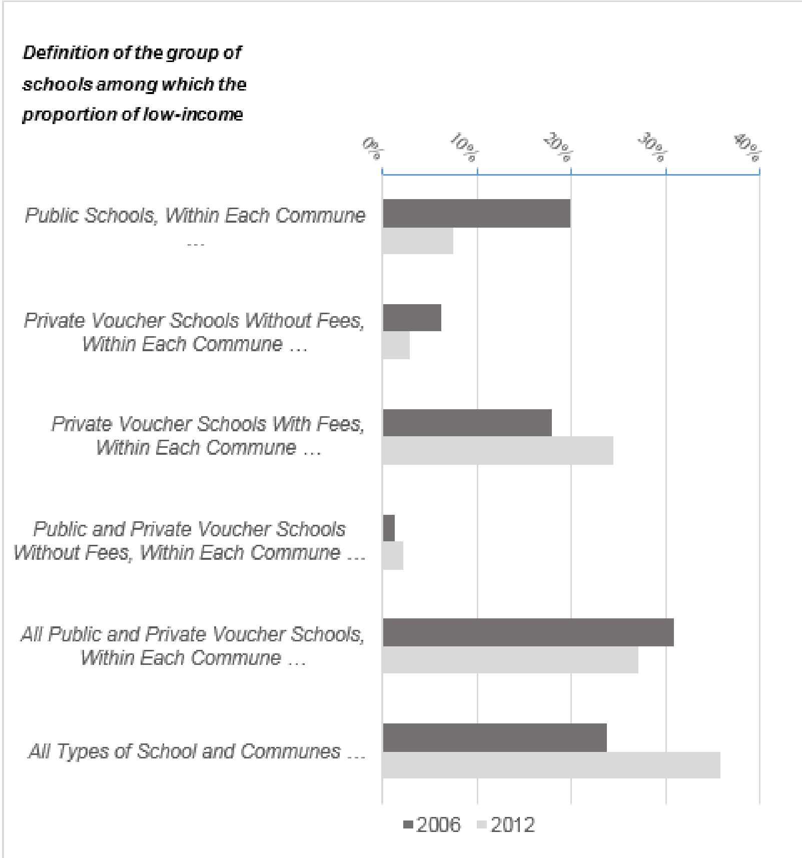

Figure 6: Display of the percentage reduction in school segregation by income that would result in 2006 (dark bars) and in 2012 (light bars) from equalizing the percentage of low-income students attending schools in a particular group.

In the next step, 4th-grade students in each commune who were enrolled either in public schools or in private voucher schools that had not charged fees are redistributed so the percentage of 4th-grade low-income students in each is equal. The net impact of this step on the amount of segregation is very small (1 percent in 2006 and 2 percent in 2012). The explanation is that public schools and private no-fee voucher schools served approximately the same percentage of 4th-grade low-income students in the two years.

Contrast this with the results of the next step, in which 4th-grade low-income students in each commune who attended public schools or any type of private voucher school are redistributed. As illustrated in Figure 6, the impact of this hypothetical redistribution on the extent of segregation is large (31 percent in 2006 and 27 percent in 2012). The explanation is that private fee-charging voucher schools served a much lower percentage of low-income 4th-grade students than did either public schools or no-fee private voucher schools. Consequently, equalizing the percentage of low-income 4th-grade students among all of these schools within each commune would reduce segregation markedly.

In the final step, 4th-grade students attending any public school or private voucher school in the country are redistributed to equalize the percentage of low-income 4th-grade students in each school. As illustrated by the bottom set of bars in Figure 6, this would reduce segregation by a larger amount in 2012 (36 percent) than in 2006 (24 percent). The explanation for this pattern is that residential segregation by income was greater in 2012 than in 2006. This reduced the potential for children from low-income families to attend the same schools as children from higher-income families.

In summary, the extent of school segregation by income among Chilean 4th-grade students was about the same in 2012, five years after the introduction of SEP, as it was in 2006. However, the pattern of segregation was quite different. In 2012, much more of the school segregation stemmed from residential segregation than was the case six years earlier. A corollary is that, in 2012, there was less school segregation by income among 4th-grade children living in the same small geographical area than was the case in 2006. One aspect of the change was that public schools in each commune were more socio-economically integrated in 2012 than they were in 2006. This occurred during a period in which the percentage of higher-income 4th-grade students attending public schools declined from 29 to 24. The explanation is that the higher-income 4th-grade students that were enrolled in public schools in each commune were more evenly distributed among the public schools in that commune in 2012 than in 2006.

Two notable patterns concern the distribution of 4th-grade students enrolled in private schools that had charged fees in 2006. The first is that the percentage of 4th-grade students in these schools who came from low-income families was higher in 2012 (41 percent) than in 2006 (33 percent). The second is that fee-charging private voucher schools were more socio-economically segregated in 2012 than they were in 2006. Some set low fees and attracted low-income students who brought with them higher-valued vouchers. Others charged higher fees and specialized in serving higher-income students.

V. Threats to Validity

Events Concurrent with SEP

Given the discontinuity design of our research, one critical threat to the validity of attributing the increase in student mathematics scores to the implementation of the SEP program are concurrent changes in the circumstances of the students and the schools they attended. Indeed, there were events that affected a great many Chilean families in the years shortly after the introduction of the SEP program. One was a sharp economic decline that occurred in 2009, following the onset of the world-wide Great Recession. A second was a series of earthquakes, including an especially devastating one that occurred in February 2010. The limited available evidence indicates that these events had negative effects on student achievement.[18] Consequently, it is unlikely that these events contributed to the improvement in student test scores during the five years after the passage of SEP.

A second threat to causal inference comes from other educational reforms that were implemented around the same time as the SEP program. For example, legislation passed in 1996 increased the length of the school day, eliminating the potential to use the same school building to educate one group of students in the morning and another in the afternoon (Bellei, 2009). While adopted well before the passage of SEP, the consequent need to build additional schools meant that the period of implementation of this new legislation was long in many areas. The additional funds provided by SEP may have enabled schools to make better use of the longer school day.[19] Using our data, it is not possible to isolate the impact of SEP implementation from those of other educational reforms. So the most defensible conclusion is that educational reforms in Chile, of which SEP was one critical part, produced substantial increases in student test scores and declines in income-based test-score gaps.

Did the cognitive skills of low-income students really improve?

In a recent paper entitled "Illusory Gains from Chile's Targeted School Voucher Experiment," Feigenberg, Rifkin, and Yan (FRY, 2017) raise doubts about the extent to which SEP closed the gap between the cognitive skills of low- and high-SES students. Based on the results of analyses of data very similar to those we use, the authors reach three conclusions:

---

18 Ananat, Gassman-Pines, & Gibson-Davis (2011) show that economic downturns had negative impacts on the academic achievement of elementary-school children in North Carolina. Gomez & Yoshikawa (forthcoming) find that exposure to the 2010 earthquake had a negative impact on the cognitive skills of young children in Chile.

19 We thank Cristian Bellei for this suggestion.

a. The gap between the average test scores of low- and high-SES students closed much less after 2008 than other studies have reported when estimated within a model that accounts for the influences of family income and parental educational attainments;

b. SEP is not responsible for increases in the relative test scores of low-SES students;

c. Increases in the relative test scores of low-SES students do not reflect real improvements in cognitive skills.

We do not disagree with the evidence these authors present. However, we do disagree with the interpretation of some of the evidence and with the conclusions they reach. We consider each conclusion in turn.

Smaller test score gap. FRY (2017) fit difference-in differences models to examine whether the gap between the mathematics and reading scores of low-SES students and higher-SES students was smaller in the years after the passage of SEP than in previous years. Similar to the results of other studies, they conclude that the size of the test-score gap declined by about 0.2 standard deviations after SEP was introduced. However, these authors go on to show (p. 13) that the test score gap closes by a much smaller amount when estimated within the context of a model that includes as covariates the family income and parental educational attainments of individual students. This is not surprising since these additional variables are indicators of the parental resources that contribute to the development of children's skills. In effect, including these family background covariates in the model controls for many of the factors that contribute to the relatively low-test scores of low-SES students. Moreover, an implicit assumption underlying the difference-in-difference models that FRY estimate is that SEP did not influence the test scores of high-SES students. It is unlikely that this assumption is valid since schools could use SEP funds to improve the education of all students. Indeed, as illustrated in Figure 1 of our paper, the average mathematics score of students at every family income percentile was higher in 2012 than in 2005.

SEP not responsible for test-score gains. FRY (2017) present several pieces of evidence in support of their conclusion that SEP was not responsible for increases in the relative test scores of low-SES students after 2007. First, they show that the additional funds SEP provided to participating schools had only a modest effect on measured inputs. They report that teachers hired with SEP funds tended to be quite inexperienced, on average, and a slightly lower percentage had a college degree than the teachers who had taught low-SES students prior to SEP. Average class size fell by less than one student per grade in schools participating in SEP.

We do not see these findings as evidence that schools used SEP funds imprudently. Indeed, a theme of a substantial literature is that reducing class size beyond the primary grades and paying for experience beyond teachers' first few years in the classroom are not effective strategies for increasing student achievement (Rivkin, Hanushek, and Kain, 2005; Hanushek and Rivkin, 2010). In contrast, strategies consistent with the evidence on changes in inputs that FRY present have closed SES-based test-score gaps in other settings. These strategies focus on how resources are used rather than on which inputs are purchased (Banerjee et al., 2007; Fryer, 2014).

FRY (2017) also point out that the test score gap did not decline more in the years after 2007 for low-SES students enrolled in schools that participated in SEP than for low-SES students enrolled in non-participating schools. This evidence is consistent with SEP making a difference. Schools chose whether to participate in SEP. Those that were thriving prior to the passage of SEP may have declined participation in order to avoid the obligations that were part of the accountability provisions of SEP. Some low-SES parents were able to enroll their child in an elementary school that was thriving without SEP. However, many schools were not thriving prior to the passage of SEP and many low-SES parents were not able to enroll their child in a high-quality elementary school. It is these schools and parents that may have benefited from SEP.

Test-score gains did not mean stronger cognitive skills. FRY (2017) present two types of evidence in support of their conclusion that the gains from SEP were "illusory." First, they show that the increases in the relative scores of low-SES students on low-stakes tests taken in 8th grade were only half as large as the increases in the relative scores on the quite high-stakes grade-4 tests. Second, they show that in the first few years after the introduction of SEP, the rate of missing scores on the national grade-4 tests increased, especially among low-SES students likely to be low-scoring. This increased the average scores of those low-SES students that did take the tests. This evidence does support the authors' argument that some Chilean elementary schools responded to accountability pressure by taking actions that did not maximize the long-term learning of students. However, we do not see this as justifying the conclusion that SEP had no meaningful impact on the quality of education provided to low-income students.

Our interpretation of the evidence in our paper as well as that in the FRY (2017) paper is informed by the literature on school improvement and especially by a recent paper by Cristian Bellei and his colleagues (2015). This paper reports the results of 12 case studies of Chilean elementary schools that had improved their performance on the national reading and mathematics tests between 2002 and 2010. Bellei and his colleagues argue that the schools that they studied followed a continuum of four paths to improved SIMCE scores, from "restricted improvement" to "institutionalized educational effectiveness." Schools following the first path were initially very low-performing and had very little capacity to provide high-quality instruction. They responded to accountability pressures by focusing intently on improving grade-4 SIMCE reading and mathematics scores. Their actions included some of the practices FRY describe. Elacqua (2016) also examined responses of low-performing Chilean elementary schools to accountability pressure and reported similar responses. Instead of investing in improving the quality of instruction, these schools hired tutors to work with low-achieving students and assigned their most qualified teachers to the fourth grade, the grade level where students take the national reading and mathematics tests used in the SEP accountability system.[20]

At the other end of the continuum, schools that had institutionalized improvement invested in developing the teaching skills of its teachers, in making instruction more consistent across grade levels, and in developing a shared sense of responsibility for the learning of all students. This process took many years of work and strong leadership. It is much more likely that increases in SIMCE scores in these schools reflected increases in children's cognitive skills than is the case in the first group of schools.

We see SEP as a complex policy initiative aimed at fostering the development of schools that would provide high-quality education to all students, including those from low-income families. The voucher system it replaced had relied on competition among schools to improve educational quality. SEP explicitly acknowledged that this was not sufficient. The law included provisions to support school improvement and hold schools accountable for improving. Schools responded to SEP in a variety of ways, not all of which were constructive. This led the legislature to revise the educational reform legislation several times. For example, as of 2015, no schools that accept vouchers, even those not participating in SEP, may charge tuition to students from low-income families.

We view the responses to SEP as encouraging, especially the increases in SIMCE scores for children from all family income percentiles, and the decline in income-based test score gaps.

---

[20] Daniel Koretz (2008) has pointed out that unproductive response of some schools to test-based accountability is inevitable. Cohen and his colleagues (2014) have shown that strategic behavior that does not benefit students is especially prevalent among schools with very limited capacity to provide coherent, consistent high-quality instruction.

However, the great variation in school improvement rates in the years after SEP that we document is troubling, as is evidence that FRY (2017) present. Refining Chile's policies for supporting schools with very different capacities and for holding them accountable in a manner that elicits constructive responses is an ongoing challenge.

VI. Concluding comments

We have argued that the combination of support and accountability that SEP provided to participating schools is the primary mechanism through which the law resulted in increased test scores, especially for low-SES students. In principle, it would have been possible for the Chilean government to introduce these provisions without a system of differentiated school vouchers. This may lead some readers to ask if the choice provisions of the SEP legislation were important. We cannot answer this question definitively because we have no evidence from Chile on the responses of schools and families to a system of support and accountability without choice. However, we have presented some evidence that the choice provisions of SEP did play a constructive role. This comes from the decomposition of the differences between 2006 and 2012 in the test score distribution of low-SES students. We found that about 10 percent of the increase in the average scores of low-SES students stemmed statistically from changes in the composition of students across school types, and 90 percent stemmed from increases in the scores of students enrolled in schools of each type.

We close by returning to the ideas of two early proponents of vouchers. Milton Friedman (1962) envisioned that the use of vouchers would improve education and increase efficiency by stimulating the supply of private schools and empowering parents to find schools that were good matches for their children. Writing prior to the 1966 publication of *Equality of Educational Opportunity* (better known as the Coleman Report, Coleman et al., 1966), which provided the first nation-wide evidence on inequality of educational outcomes and school segregation by race and class, Friedman did not emphasize these concerns. Writing after the Coleman Report had received significant attention, Christopher Jencks (1970) was concerned centrally with these topics and they influenced the design of the regulated compensatory voucher system that he proposed. Chile's experience with a universal voucher system in the years before and after SEP shows that the design of regulations and incentives matter greatly.

References

Abdulkadiroğlu, Atila, Parag A. Pathak, and Christopher R. Walters (forthcoming). "Free to Choose: Can School Choice Reduce Student Achievement?" *American Economic Journal: Applied Economics.*

Ananat, Elizabeth O., Anna Gassman-Pines, and Christina M. Gibson-Davis. 2011. "The Effects of Local Employment Losses on Children's Educational Achievement." In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murnane, 299-314. New York: Russell Sage Foundation and the Spencer Foundation.

Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 122 (3): 1235-1264.

Bellei, Cristian. 2009. "Does Lengthening the School Day Increase Students' Academic Achievement? Results from a Natural Experiment in Chile." *Economics of Education Review* 28 (5): 629-640.

Bellei, Cristian and Xavier Vanni. 2015. "The Evolution of Educational Policy in Chile, 1980-2014." In *Education in South America*, edited by Simon Schwartzman. London: Bloomsbury Publishing.

Bellei, Cristian, Xavier Vanni, Juan Pablo Valenzuela, and Daniel Contreras. 2015. "School Improvement Trajectories: An Empirical Typology." *School Effectiveness and School Improvement* 10: 1-18.

Carrasco, Rafael. 2014. "Leveling the Playing Field: How can we Address Educational Inequalities?" Stanford University Graduate School of Education.

Carrell, Scott E. and Mark L. Hoekstra. 2010. "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids." *American Economic Journal: Applied Economics* 2 (1): 211-228.

Carrell, Scott E., Mark Hoekstra, and Elira Kuka. 2016. The Long-Run Effects of Disruptive Peers. Cambridge, MA: National Bureau of Economic Research Working Paper No. 22042.

Clotfelter, Charles T. 2004. *After Brown : The Rise and Retreat of School Desegregation*. Princeton, N.J.: Princeton University Press.

Cohen, David K., Donald J. Peurach, Joshua L. Glazer, Karen E. Gates, and Simona Goldin. 2014. *Improvement by Design : The Promise of Better Schools*.The University of Chicago Press.

Coleman, James Samuel, Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic Weinfeld, and Robert L. York. 1966. *Equality of Educational Opportunity*. Washington, D.C.: U.S. Govt. Print. Off

Contreras, Dante, Paulina Sepulveda, and Sebastian Bustos. 2010. "When Schools are the Ones that Choose:The Effects of Screening in Chile." *Social Science Quarterly* 91 (5): 1349-1368.

Duncan, Greg J. and Katherine Magnuson. 2011. "The Nature and Impact of Early Achievement Skills, Attention Skills, and Behavior Problems." In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murnane, 47-70. New York: Russell Sage and Spencer Foundations.

Dynarski, Mark, Ning Rui, Ann Webber, and Babette Gutmann. 2017. Evaluation of the DC Opportunity Scholarship Program: Impacts After One Year. Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Elacqua, Greg, Matias Martinez, Humberto Santos, and Daniela Urbina. 2016. "Short- Run Effects of Accountability Pressures on Teacher Policies and Practices in the Voucher System in Santiago, Chile." *School Effectiveness and School Improvement* 27 (3): 385-405.

Elacqua, Gregory. 2012. "The Impact of School Choice and Public Policy on Segregation: Evidence from Chile." *International Journal of Educational Development* 32 (3): 444-453.

Epple, Dennis and Richard E. Romano. 1998. "Competition between Private and Public Schools: Vouchers, and Peer-Group Effects." *American Economic Review* 88 (1): 33-62.

Epple, Dennis, Richard E. Romano, and Urquiola Urquiola. 2017. "School Vouchers: A Survey of the Economics Literature." *Journal of Economic Literature* 55 (2): 441-492.

Feigenberg, Benjamin, Steven Rivkin, and Rui Yan. 2017. Illusory Gains from Chile's Targeted School Voucher Experiment. Cambridge MA: National Bureau of Economic Research Working Paper 23178.

Friedman, Milton. 1962. *Capitalism and Freedom*. Chicago: University of Chicago Press.

Fryer, Roland G., Jr. 2014. "Injecting Charter School Best Practices into Traditional PublicSchools:Evidence from Field Experiments." *Quarterly Journal of Economics* 129 (3): 1355-1407.

Gomez, Celia and Hirokazu Yoshikawa (forthcoming). "Earthquake Effects: Estimating the Relationship between Exposure to the 2010 Chilean Earthquake and Preschool Children's Early Learning Outcomes." *Early Childhood Research Quarterly*.

Gonzales, Patrick, Christopher Calsyn, Leslie Jocylyn, Kitty Mak, David Kastberg, Sousan Arafeh, Trevor Williams, and Winnie Tsen. 2000. *Pursuing Excellence: Comparisons of International Eighth-Grade Mathematics and Science Achievement from a U.S. Perspective, 1995 and 1999*. Washington, DC: U.S. Department of Education. National Center for Education Statistics.

Hanushek, Eric A. and Steven G. Rivkin. 2010. "Generalizations about using Value- Added Measures of Teacher Quality." *American Economic Review* 100 (2): 267-271.

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.

Jackson, C. Kirabo, Rucker C. Johnson, and Claudia Persico. 2016. "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms." *Quarterly Journal of Economics* 131 (1): 157-218.

Jencks, Christopher. 1970. *Education Vouchers: A Report on Financing Education by Payments to Parents*. Cambridge MA: Center for the Study of Public Policy.

Koretz, Daniel M. 2008. *Measuring Up: What Educational Testing really Tells Us*. Cambridge, Mass. : Harvard University Press.

Murnane, Richard J., John B. Willett, and Frank Levy. 1995. "The Growing Importance of Cognitive Skills in Wage Determination." *Review of Economics and Statistics* 77 (2): 251-266.

Navarro-Palau, Patricia. 2016. "Effects of Differentiated School Vouchers: Evidence from a Policy Change and Date of Birth Cutoffs." Columbia University.

Nechyba, Thomas J. 2006. "Income and Peer Quality Sorting in Public and Private Schools." In *Handbook of the Economics of Education,* edited by Eric Alan Hanushek and Finis Welch, 1327-1368. Amsterdam ; London: North-Holland.

———. 2000. "Mobility, Targeting, and Private-School Vouchers." *American Economic Review* 90 (1): 130-146.

Neilson, Christopher. 2015. "Targeted Vouchers, Competition among Schools, and the Academic Achievement of Poor Students." Yale University.

Oaxaca, Ronald L. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14 (3): 693-709.

Reardon, Sean F. 2011. "The Widening Academic Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations." In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murnane, 91-116. New York: Russell Sage Foundation and the Spencer Foundation.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73 (2): 417-458.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Valenzuela, Juan Pablo, Cristian Bellei, and Danae de los Rios. 2013. "Socioeconomic School Segregation in a Market-Oriented Educational System. the Case of Chile." *Journal of Education Policy* 29 (2): 217-241

Appendix A

Detailed Specifications of the Multilevel Models, By Research Question

RQ1: Trends in Student Achievement Over Time

To address our first research question (*RQ1*), we specified students' raw mathematics scores as a function of: (a) the passage of time (*YEAR*), (b) the onset of the *SEP* policy implementation (*SEP*), and (c) family-income percentile (***INC***), in a multilevel statistical model. For the k^{th} fourth-grade student enrolled in the i^{th} school in the j^{th} year, the multilevel (*Level-1/Level-2*) specification of our full model for addressing *RQ1* is:

Level-1/Student-Year:

$$
\begin{aligned}
MATH_{ijk} = \pi_{0i} &+ \pi_{1i}(YEAR_j - 2008) \\
&+ \pi_{2i}SEP_j + \pi_{3i}[SEP_j \times (YEAR_j - 2008)] \\
&+ \boldsymbol{\beta}_1'\boldsymbol{INC}_{ijk} + \boldsymbol{\beta}_2'[\boldsymbol{INC}_{ijk} \times (YEAR_j - 2008)] \\
&+ \boldsymbol{\beta}_3'(\boldsymbol{INC}_{ijk} \times SEP_j) + \boldsymbol{\beta}_4'[(\boldsymbol{INC}_{ijk} \times SEP_j) \times (YEAR_j - 2008)] \\
&+ \{\epsilon_{ijk} + \delta_{ij}\}
\end{aligned}
$$

where $\epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$ and $\delta_{ij} \sim N(0, \sigma_\delta^2)$ [A1]

***Level-2/School*:**

$$\pi_{0i} = \gamma_{00} + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \zeta_{1i}$$

$$\pi_{2i} = \gamma_{20} + \zeta_{2i}$$

$$\pi_{3i} = \gamma_{30} + \zeta_{3i}$$

where $\begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \\ \zeta_{2i} \\ \zeta_{3i} \end{bmatrix} \sim MVN_4(\boldsymbol{0}, \boldsymbol{\Sigma}_\zeta)$

Notice that, at Level-1 -- the *Student/Year* level, in the fixed part of the model, we have included terms to represent the main effect of time-varying predictors *YEAR* and *SEP*, and their two-way interaction. This part of the model specification accounts for the standard features of our discontinuity design and permits the estimation of unique population average trends in achievement over time in the pre- and post-SEP periods, by school. Then, to allow these trajectories to differ by family income, we have also included the main effect of predictor vector $\boldsymbol{INC}$ (containing the student's family-income percentile, its square, and its cube). The model also contains cross-products representing the two-way interactions between family-income percentile and time, and between family-income percentile and *SEP*, plus the three-way interaction among all three predictors. These latter components of the specification permit the hypothesized relationship between student mathematics achievement and both time and family income to differ in the pre- and post-SEP periods. To reduce both model complexity and computing burden (which was extreme), we have fixed -- across schools -- the effects of the predictors that represented family-income percentile. Finally, in addressing *RQ1*, we have included no predictors at the school level, but have simply permitted the corresponding *Level-1* parameters to differ around their model-specified population averages (γ_{00}, γ_{10}, γ_{30}, γ_{40}), with the corresponding hypothesized population variances.

In the multilevel model in [A1], we have accounted for the complex levels of nesting present in our data-design by including selected random effects at each level. At *Level-1*, we include two random effects. First, we have hypothesized that -- in the population – residuals ϵ_{ijk} are distributed independently and identically normal with mean zero and variance σ_{ϵ}^2. Second, because students are nested within schools and a single testing instance (that is, year), we have included the random effect of school and year, δ_{ij}, again assumed to be distributed independently and identically normal, but with variance σ_{δ}^2. At the school level, because each school contributes multiple years of testing data, we have hypothesized that the school-level achievement trajectories possess a random intercept and a random slope across schools, denoted by π_{0i} and π_{1i} in the pre-SEP period, along with increments to both, denoted by π_{2i} and π_{3i} respectively in the post-SEP period. Finally, we have assumed that these latter four school-level random effects are distributed multivariate normal with mean vector zero and unconstrained covariance matrix $\boldsymbol{\Sigma}_{\zeta}$.

Of course, one need not rely solely on a multilevel specification. All such models can be collapsed algebraically into a corresponding composite model, which has the appearance of a standard linear statistical (regression) model, but incorporates a complex error term to account for the nested and time-varying nature of the data design. For *RQ1*, for instance, substituting from *Level-2* of the specification into *Level-1*, this composite model becomes:

$$\begin{aligned} MATH_{ijk} = \gamma_{00} + \gamma_{10}(YEAR_j - 2008) \\ +\gamma_{20}SEP_j + \gamma_{30}[SEP_j \times (YEAR_j - 2008)] \\ +\boldsymbol{\beta}_1'\boldsymbol{INC}_{ijk} + \boldsymbol{\beta}_2'[\boldsymbol{INC}_{ijk} \times (YEAR_j - 2008)] \\ +\boldsymbol{\beta}_3'(\boldsymbol{INC}_{ijk} \times SEP_j) + \boldsymbol{\beta}_4'[(\boldsymbol{INC}_{ijk} \times SEP_j) \times (YEAR_j - 2008)] \\ +\{(\epsilon_{ijk} + \delta_{ij}) + (\zeta_{0i} + \zeta_{1i}(YEAR_j - 2008) + \zeta_{2i}SEP_j \\ + \zeta_{3i}[SEP_j \times (YEAR_j - 2008)])\} \end{aligned} \qquad [[A2]$$

Or, more simply:

$$\begin{aligned} MATH_{ijk} = \gamma_{00} + \gamma_{10}(YEAR_j - 2008) \\ +\gamma_{20}SEP_j + \gamma_{30}[SEP_j \times (YEAR_j - 2008)] \\ +\boldsymbol{\beta}_1'\boldsymbol{INC}_{ijk} + \boldsymbol{\beta}_2'[\boldsymbol{INC}_{ijk} \times (YEAR_j - 2008)] \\ +\boldsymbol{\beta}_3'(\boldsymbol{INC}_{ijk} \times SEP_j) + \boldsymbol{\beta}_4'[(\boldsymbol{INC}_{ijk} \times SEP_j) \times (YEAR_j - 2008)] + \epsilon_{ijk}' \end{aligned} \qquad [[A3]$$

Where ϵ_{ijk}' is a composite multilevel time-dependent residual, given by:

$$\epsilon_{ijk}' = \left\{(\epsilon_{ijk} + \delta_{ij}) + \begin{pmatrix} \zeta_{0i} + \zeta_{1i}(YEAR_j - 2008) + \zeta_{2i}SEP_j \\ +\zeta_{3i}[SEP_j \times (YEAR_j - 2008)] \end{pmatrix}\right\} \qquad [[A4]$$

With constituent random effects distributed as assumed in [A1]. Inspecting [A4] confirms that – while our specification describes the complex nested structure of our data – it permits the hypothesized level-2 error-covariance structure (representing the population variances and covariances among trends in student achievement over time) to be heteroscedastic across the pre- and post-SEP periods. It is the composite multilevel model in [A3] that is listed in the text as the principal model whose fitting permits us to address *RQ1*.

RQ2: The Impact of SEP on Student Achievement Trends

In addressing *RQ2*, we sought to discern differences in the impact of the implementation of the SEP policy on student mathematics achievement among different types of schools, as distinguished by their location and organizational type. In doing so, we regarded it critical to control for the effects of selected important student and family characteristics (student gender, whether a student had repeated fourth grade, family income, and parental educational attainment). This is because it may have been easier for schools to improve their average student achievement post-SEP by attracting more advantaged students than by improving the intrinsic quality of the education they provided. However, we also wanted to recognize that an explicit goal of SEP implementation itself was to reduce achievement gaps based on socioeconomic status. If SEP succeeded, we anticipated that parameters associated with these same selected student and family covariates in our statistical models would have smaller values in the post-SEP period than in the pre-SEP. In order to control for student and family background influences, but also attribute to SEP reductions in the influences of these variables on student achievement, we constrained the parameters on the selected covariates to those values they had during the three-year period prior to SEP implementation.

Unfortunately, we could not impose the required constraints directly during the fitting of subsequent statistical models due to limitations in our model-fitting software and the size of our dataset -- limitations that were exacerbated by the multilevel nature of our data and our concurrent implementation of the methods of multiple imputation. So, in advance of any analyses to address *RQ2*, we chose to adjust the values of our mathematics outcome by *partialling* the effects of the selected covariates from it. We did this using a two-step procedure. First, using only data on fourth-grade students during the pre-SEP years (2005-2007), we fitted a statistical model to predict student mathematics score as a function of the selected student and family background covariates, and estimated their associated slope parameters. The estimated parameters of this model are listed below in Appendix A, Table A1. Then, using these estimated parameter values, we partialled the effect of the selected covariates from the original *MATH* outcome and constructed a measure of *adjusted* student mathematics achievement, *ADJ_MATH*, for every student, in both the pre-SEP and post-SEP periods. All subsequent analyses to address *RQ2* treated *ADJ_MATH* – rather than *MATH* -- as the outcome.

To address *RQ2*, we amended our multilevel specification in [A1], for the k^{th} fourth-grade student enrolled in the i^{th} school in the j^{th} year, as follows:

Level-1/Student-Year:

$$ADJ\_MATH_{ijk} = \pi_{0i} + \pi_{1i}(YEAR_j - 2008) + \pi_{2i}SEP_j + \pi_{3i}\{SEP_j \times (YEAR_j - 2008)\} + \{\epsilon_{ijk} + \delta_{ij}\}$$

where $\epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$ and $\delta_{ij} \sim N(0, \sigma_\delta^2)$

***Level-2/School*:** [[A5]

$$\pi_{0i} = \gamma_{00} + \boldsymbol{\gamma_o'W_i} + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \boldsymbol{\gamma_1'W_i} + \zeta_{1i}$$

$$\pi_{2i} = \gamma_{20} + \boldsymbol{\gamma_2'W_i} + \zeta_{2i}$$

$$\pi_{3i} = \gamma_{30} + \boldsymbol{\gamma_3'W_i} + \zeta_{3i}$$

where $\begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \\ \zeta_{2i} \\ \zeta_{3i} \end{bmatrix} \sim MVN_4(\boldsymbol{0}, \boldsymbol{\Sigma}_{Adj\zeta})$

Notice that, at *Level-1* (*Student/Year*) of our multilevel model in [A5], we have again included time-varying predictors to capture the main effect of both *YEAR* and *SEP*, and their two-way interaction. This part of the model specification accounts for the standard features of our discontinuity design. As before, this specification permits unique population average trends in achievement over time in the pre- and post-SEP periods, by school. Notice, though, that we have eliminated the direct effects of family-income percentile in the model because this predictor has already been partialled from the student mathematics achievement score during the outcome adjustment process described above. In our *Level-2* school-level model, we have now added the time-invariant predictor vector $\boldsymbol{W}$, to distinguish among schools based on their organizational form and location. It is the effects of these latter predictors that address our second research question. In the multilevel model, we have again accounted for the complex levels of nesting present in our data-design by including selected random effects at each level, similar to those hypothesized under *RQ1*, in [A1] above.

Again, the specified multilevel model can be collapsed into a corresponding composite model, with a complex error term that accounts for the nested and time-varying nature of the data design. For *RQ2*, substituting from *Level-2* of the specification into *Level-1*, this composite model is:

$$
\begin{aligned}
ADJ\_MATH_{ijk} = {} & \gamma_{00} + \boldsymbol{\gamma_o'} \boldsymbol{W_i} + \gamma_{10}(YEAR_j - 2008) \\
& + \boldsymbol{\gamma_1'}[\boldsymbol{W_i} \times (YEAR_j - 2008)] + \gamma_{20} SEP_j + \boldsymbol{\gamma_2'}[\boldsymbol{W_i} \times SEP_j] \\
& + \gamma_{30}\{SEP_j \times (YEAR_j - 2008)\} \\
& + \boldsymbol{\gamma_3'}[\boldsymbol{W_i} \times SEP_j \times (YEAR_j - 2008)] \\
& + \{(\epsilon_{ijk} + \delta_{ij}) \\
& \quad + (\zeta_{0i} + \zeta_{1i}(YEAR_j - 2008) + \zeta_{2i} SEP_j \\
& \quad + \zeta_{3i}[SEP_j \times (YEAR_j - 2008)])\}
\end{aligned}
\qquad \text{[[A6]}
$$

Or, more simply:

$$
\begin{aligned}
ADJ\_MATH_{ijk} = {} & \gamma_{00} + \boldsymbol{\gamma_o'} \boldsymbol{W_i} + \gamma_{10}(YEAR_j - 2008) \\
& + \boldsymbol{\gamma_1'}[\boldsymbol{W_i} \times (YEAR_j - 2008)] + \gamma_{20} SEP_j + \boldsymbol{\gamma_2'}[\boldsymbol{W_i} \times SEP_j] \\
& + \gamma_{30}\{SEP_j \times (YEAR_j - 2008)\} \\
& + \boldsymbol{\gamma_3'}[\boldsymbol{W_i} \times SEP_j \times (YEAR_j - 2008)] + \boldsymbol{\epsilon_{ijk}''}
\end{aligned}
\qquad \text{[[A7]}
$$

Where ϵ_{ijk}'' is a composite multilevel time-dependent residual, given by:

$$
\epsilon_{ijk}'' = (\epsilon_{ijk} + \delta_{ij}) + \begin{pmatrix} \zeta_{0i} + \zeta_{1i}(YEAR_j - 2008) + \zeta_{2i} SEP_j \\ + \zeta_{3i}[SEP_j \times (YEAR_j - 2008)] \end{pmatrix} \qquad \text{[A8]}
$$

With constituent random effects distributed as assumed in [A5] above. Inspecting [A8] confirms that – while our specification accounts for the complex nested structure of our data – it permits the hypothesized level-2 error-covariance structure (representing the population variances and covariances among trends in student achievement over time) to be heteroscedastic across the pre- and post-SEP periods. It is composite multilevel model [A7] that is listed in the body of the text as the principal model whose fitting permits us to address *RQ2*.

Appendix A Table A1. Estimates, standard errors and approximate *p*-values from a fitted multilevel model summari relationship between fourth-grade student mathematics scores, in Chile, and: (a) year (re-parameterized as dichotom predictors, representing 2005 thru 2007, with 2005 omitted), (b) a cubic polynomial function of family-income perc (c) parental educational attainment (re-parameterized as a vector of dichotomous predictors, uniquely for both moth father, with lowest category omitted, in each case), (d) school location, and (e) student gender and repeater status. estimates were obtained using the method of multiple-imputation to account for the presence of missing data (m=8)

Effects	**Model Parameters**	**Parameter Estimate**	**Standard Error**
Intercept	γ_{00}	233.887***	0.431
YEAR06	γ_{10}	0.256	0.305
YEAR07	γ_{20}	-2.325***	0.321
RURAL	γ_{01}	-11.834***	0.732
YEAR06×RURAL	γ_{11}	-0.443	0.646
YEAR07×RURAL	γ_{21}	1.655*	0.667
INC	β_1	8.650***	0.641
INC^2	β_2	-6.869***	1.091
INC^3	β_3	30.527***	4.437
FEMALE	θ_1	-5.635***	0.130
FEMALE×RURAL	θ_2	4.344***	0.388
REPEATER	φ_1	-47.100***	0.338
REPEATER×RURAL	φ_2	7.499***	1.007
PA_ED2	ρ_{12}	2.529***	0.290
PA_ED3	ρ_{13}	3.359***	0.252
PA_ED4	ρ_{14}	7.083***	0.261
PA_ED5	ρ_{15}	10.796***	0.317
MA_ED2	τ_{12}	4.077***	0.281
MA_ED3	τ_{13}	5.203***	0.254
MA_ED4	τ_{14}	12.002***	0.272
MA_ED5	τ_{15}	15.614***	0.301
PA_ED2×RURAL	ρ_{22}	2.218***	0.593
PA _ED3×RURAL	ρ_{23}	1.762*	0.720
PA _ED4×RURAL	ρ_{24}	1.592*	0.692
PA _ED5×RURAL	ρ_{25}	0.945	1.132
MA_ED2×RURAL	τ_{22}	1.906***	0.591
MA_ED3×RURAL	τ_{23}	3.242***	0.650
MA_ED4×RURAL	τ_{24}	3.832***	0.674
MA_ED5×RURAL	τ_{25}	6.238***	1.098
***Random Effects*:**			
Level-1:			
Student	σ_{ε}^2	2203.277	
Year	σ_{δ}^2	79.018	
Level-2:			
School	$\sigma_{\zeta 0}^2$	347.058	
	$\sigma_{\zeta 1}^2$	75.201	
	$\sigma_{\zeta 2}^2$	112.687	
	$\sigma_{\zeta 0 \zeta 1}$	-18.100	
	$\sigma_{\zeta 1 \zeta 2}$	42.267	
	$\sigma_{\zeta 0 \zeta 2}$	-23.828	
***Goodness-of-Fit and Associated Statistics*:**			
Model *F*-Statistic		1482.47***	
Number of Students		646,979	
Number of Schools		7,968	

Key: * *p*<.05, ** *p*<.01, *** *p*<.001

Appendix B Table B1: Estimates, standard errors and approximate p-values from a parsimonious fitted multilevel model that summarizes the relationship between students' fourth-grade mathematics scores, in Chile, and: (a) chronological year (2005 thru 2012), (b) the implementation of the SEP program and (c) family-income percentile. All estimates obtained using the method of multiple-imputation to account for the presence of missing data (m=8).

Effects	Model Parameters	Parameter Estimate	Standard Error
***Fixed Effects*:**			
Intercept	γ_{00}	234.85***	0.42
(*YEAR*-2008)	γ_{10}	-1.66***	0.14
SEP	γ_{20}	1.10**	0.35
SEP×(*YEAR*-2008)	γ_{30}	5.42***	0.16
INC	β_{11}	13.05***	0.79
INC^2	β_{12}	-9.14***	1.01
INC^3	β_{13}	42.16***	2.45
INC×(*YEAR*-2008)	β_{21}	-3.82***	0.33
INC×*SEP*	β_{31}	6.17***	0.81
INC^2×*SEP*	β_{32}	12.66***	1.30
INC×*SEP*×(*YEAR*-2008)	β_{41}	2.48***	0.37
***Random Effects*:**			
Student/Year Level:			
	σ_{ε}^2	2129.46***	
	σ_{δ}^2	127.91***	
School-Level:			
	$\sigma_{\zeta 0}^2$	574.76***	
	$\sigma_{\zeta 1}^2$	6.03***	
	$\sigma_{\zeta 2}^2$	10.12	
	$\sigma_{\zeta 3}^2$	19.78***	
	$\sigma_{\zeta 0 \zeta 1}$	37.54***	
	$\sigma_{\zeta 0 \zeta 2}$	-35.36***	
	$\sigma_{\zeta 0 \zeta 3}$	-74.09***	
	$\sigma_{\zeta 1 \zeta 2}$	0.42	
	$\sigma_{\zeta 1 \zeta 3}$	-6.76***	
	$\sigma_{\zeta 2 \zeta 3}$	1.51	
Goodness-of-Fit and Associated Statistics:			
Model F-Statistic		2108.71***	
Number of Students		1,631,841	
Number of Schools		8,464	

Key: * p<.05, ** p<.01, *** p<.001

Appendix B Table B2: Estimates, standard errors and approximate *p*-values from a fitted multilevel model summarizing the relationship between fourth-grade student adjusted-mathematics scores, in Chile, and: (a) YEAR (centered on 2008), (b) implementation of *SEP* and its interaction with centered *YEAR*, (c) school location, and (d) school organizational type. All estimates were obtained using the method of multiple-imputation to account for the presence of missing data (m=8).

Effects	**Model Parameters**	**Parameter Estimate**	**Standard Error**
***Fixed Effects*:**			
INTERCEPT	γ_{00}	232.421***	0.630
RURAL	γ_{01}	-2.034***	0.915
NOFEE_NFP	γ_{02}	12.360***	1.666
NOFEE_FP	γ_{03}	4.349**	1.253
FEE_NFP	γ_{04}	22.865***	1.434
FEE_FP	γ_{05}	15.140***	1.052
RURAL×NOFEE_NFP	γ_{06}	-17.906***	1.923
RURAL×NOFEE_FP	γ_{07}	-12.595***	1.105
(YEAR-2008)	γ_{10}	-2.034***	0.224
(YEAR-2008)×RURAL	γ_{11}	0.275	0.352
(YEAR-2008) ×NOFEE_NFP	γ_{12}	0.678	0.630
(YEAR-2008) ×NOFEE_FP	γ_{13}	3.256***	0.486
(YEAR-2008) ×FEE_NFP	γ_{14}	1.059*	0.524
(YEAR-2008) ×FEE_FP	γ_{15}	1.757***	0.392
SEP	γ_{20}	1.698**	0.536
SEP×RURAL	γ_{21}	1.596	0.847
SEP×NOFEE_NFP	γ_{22}	-0.055	1.493
SEP×NOFEE_FP	γ_{23}	-3.254**	1.138
SEP×FEE_NFP	γ_{24}	0.188	1.233
SEP×FEE_FP	γ_{25}	-2.367*	0.923
(YEAR-2008)×SEP	γ_{30}	6.603***	0.257
(YEAR-2008)×SEP×RURAL	γ_{31}	-0.902*	0.404
(YEAR-2008)×SEP×NOFEE_NFP	γ_{32}	-1.481*	0.711
(YEAR-2008)×SEP×NOFEE_FP	γ_{33}	-2.875***	0.547
(YEAR-2008)×SEP×FEE_NFP	γ_{34}	-4.041***	0.592
(YEAR-2008)×SEP×FEE_FP	γ_{35}	-3.846***	0.441
***Random Effects*:**			
Level-1:			
Student	σ_ε^2	2045.464	
Year	σ_δ^2	128.684	
Level-1: School			
	$\sigma_{\zeta 0}^2$	356.119	
	$\sigma_{\zeta 1}^2$	7.154	
	$\sigma_{\zeta 2}^2$	17.039	
	$\sigma_{\zeta 3}^2$	18.158	
	$\sigma_{\zeta 0 \zeta 1}$	27.657	
	$\sigma_{\zeta 0 \zeta 2}$	-25.743	
	$\sigma_{\zeta 0 \zeta 3}$	-49.084	
	$\sigma_{\zeta 1 \zeta 2}$	-2.845	
	$\sigma_{\zeta 1 \zeta 3}$	-7.481	
	$\sigma_{\zeta 2 \zeta 3}$	4.606	
Goodness-of-Fit and Associated Statistics:			
Model *F*-Statistic		200.64***	
Number of Students		1,631,841	
Number of Schools		8,464	

Key: * *p*<.05, ** *p*<.01, *** *p*<.001