PATRICK J. MCEWAN
MIGUEL URQUIOLA
EMILIANA VEGAS

School Choice, Stratification, and Information on School Performance: Lessons from Chile

In the early 1980s, Chile implemented a nationwide school choice system, under which the government finances education via a flat per-student subsidy (or voucher) to the public or private school chosen by a family. At present, about 94 percent of all schools (public, religious, and secular private) are voucher funded. More than half of urban schools are private, and most of these operate as for-profit institutions.[1] Since the early 1990s, Chile has also publicized information on school performance and increased per pupil expenditure substantially.

Despite these and other reforms, Chile has found it challenging to improve students' learning outcomes.[2] Hsieh and Urquiola find that the country's relative performance in international tests did not change much between 1970 and 1999.[3] Its performance on the 2000 Programme for International Student Assessment (PISA) is not only much lower than the OECD average but is similar to that of other Latin American countries and low relative to countries with similar income per capita.[4]

McEwan is with Wellesley College; Urquiola is with Columbia University; Vegas is with the World Bank.

For useful conversations and comments, we thank Eduardo Engel, Francisco Ferreira, Francisco Gallego, Alejandra Mizala, and Reynaldo Fernandes. For generous funding, we thank the Institute for Social and Economic Research and Policy (ISERP) at Columbia University. Urquiola is also grateful for support from a National Academy of Education/Spencer Foundation Postdoctoral Fellowship.

1. Elacqua (2006).
2. In much of Latin America—not just Chile—the rapid progress in terms of quantity of education has not been matched by increases in quality, as measured, for example, by test scores (Hanushek and Woessman 2007; Vegas and Petrow 2007).
3. Hsieh and Urquiola (2006).
4. Vegas and Petrow (2007).

Recent discussions and events in Chile reflect disappointment consistent with such findings. For instance, Engel and Navia observe that Chile is doing relatively well in terms of education quantity, but confronts problems with respect to quality.[5] In 2006, student protests resulted in renewed government commitments to address education quality. Additionally, there was surprising agreement among candidates in the last presidential election that education policy should address high levels of inequality.

Chile's experience suggests that we have much to learn about the consequences of education reform, especially school choice, and why it has not resulted in learning gains of the magnitude one might have predicted. This paper addresses these issues in two ways. First, we review the previous literature on the impact of choice in Chile, focusing on the effects on average student achievement and stratification. The reform was implemented nationwide, without an experimental design. Naturally, identifying its causal effect is very difficult and the literature has yet to reach a consensus in this area. To shed additional light on this debate, we present new evidence from a regression-discontinuity design that supports previous findings that school choice—at least as it has been implemented until now—has increased stratification while having little effect on average achievement.

Second, we explore some factors that may explain why the evolution of school quality in Chile has been disappointing. We focus on the difficulties researchers have encountered in generating and interpreting information on school performance. Chile has been at the forefront of measuring, analyzing, and disseminating data on student achievement. Our research suggests three cautionary tales regarding such efforts. First, schools' average test scores are a very good proxy for average student income in Chile, so they are of limited use to either parents or policymakers in terms of identifying especially effective or high value added schools. Second, commonsense approaches to control for parent and student characteristics—such as analyzing school gains between years—yield volatile school rankings. That is, assessments about whether a given school is "good" or "bad" can change substantially between years, misleading parents and policymakers. Third, the presence of test-score volatility complicates the evaluation of education programs that are assigned on the basis of test scores. We show how this volatility in test scores produced misleading estimates of the impact of one of Chile's signature education programs in the early 1990s.

5. Engel and Navia (2006). The authors further note that there is little evidence of improvement either in Chile's own SIMCE tests or in the 1999 and 2003 rounds of the Third International Mathematics and Science Study (TIMSS).

We conclude by discussing the implications of our findings for current education policy in Chile. While we touch on multiple issues, we highlight three conclusions. First, over the past three decades, Chile's school system has improved its users' welfare along many dimensions. At the very least, choice is likely to have raised the welfare of some households by letting them attend schools of their preference. The evolution of aggregate learning outcomes has not been as satisfactory, however. Second, despite concerns regarding the use of test score information for accountability purposes, we argue that the Chilean government should continue to collect and improve the use of student performance data. Third, Chile would be in a position to learn much more if it were willing to credibly evaluate the myriad education initiatives it has introduced.

The rest of the paper proceeds as follows. The next section provides background on Chile's education system and describes the national student assessment system data. We then review evidence on the impact of Chile's school choice scheme on stratification and student achievement. The paper next discusses the challenges of using test score data to assess and improve education quality. Finally, we discuss policy implications.

Background on Chile's Education System

Prior to the 1981 education reforms, three types of schools existed in Chile: fiscal or public schools were managed by the national Ministry of Education and accounted for about 80 percent of enrollments; unsubsidized private schools did not receive public funding, catered primarily to upper-income households, and accounted for about 6 percent of enrollments; and subsidized private schools did not charge tuition, received limited lump-sum public subsidies, were often religious (mostly Catholic), and accounted for roughly 14 percent of enrollment. This basic structure remained after the 1981 education reform, which had three main components. First, the ministry transferred fiscal school management to more than 300 municipalities, which began to receive a per-student subsidy for every child attending their schools. These schools retained their role as suppliers of last resort: they are not allowed to charge tuition, and they cannot turn away students unless they are over-enrolled.[6] Second, subsidized (or voucher) private schools began to receive

6. Primary public schools cannot charge tuition; as of 1994, secondary schools can, but few actually do.

the same per-student subsidy as municipal schools. Unlike the latter, they have wide latitude regarding student selection and, as of 1994, are allowed to charge fees.[7] Third, unsubsidized, tuition-charging private schools have continued to operate without public funding. While they could have started to accept vouchers, the majority of these elite institutions chose not to do so.

By 2003, the two types of private institutions together accounted for about 45 percent of all schools, while voucher schools alone accounted for about 36 percent. In urban areas, these shares were 62 and 48 percent, respectively. As stated, private voucher schools can operate as for-profits, and Elacqua calculates that about 74 percent do so.[8]

Parent-reported data in table 1 confirm substantial differences in the income and fee payments of families in municipal and private schools, especially when compared to the six percent of students enrolled in unsubsidized private sector schools (the data are from a census of fourth grade students). The data also confirm anecdotal impressions and prior research, which suggest that private schools are more likely to subject parents to admissions requirements like exams, interviews, and requests for marriage or baptismal certificates.[9]

In 1988, Chile introduced a national student assessment system known as the *Sistema Nacional de Medición de la Calidad de la Educación,* or SIMCE (the Ministry of Education also administered tests between 1982 and 1984, under a different system). Three features of these data are relevant to our subsequent discussions. First, the SIMCE assesses all students in a single grade—namely, fourth, eighth, or tenth—in a given year. Second, student-level scores and socioeconomic characteristics are only available since 1997, when the Ministry began applying a detailed questionnaire to parents. Previously, test-score data were reported as school averages.[10] Third, the scores of individual students cannot presently be linked across time to form a true panel, which would permit further analyses of school and teacher effectiveness. In the next two sections, we review evidence on the effects of Chile's reforms, all of it using one or more years of the SIMCE assessments.

7. When private subsidized schools charge fees, the per-student subsidy is adjusted downward.

8. Elacqua (2006).

9. Gauri (1998).

10. Student-level data are available for a few years before 1997 but without corresponding information on students or families.

TABLE 1. Fourth Grade Parent Characteristics in Municipal, Subsidized Private, and Unsubsidized Private Schools, 2005[a]
Percent

Characteristic	*Municipal*	*Subsidized private*	*Unsubsidized private*
Percent of total students	50.2	43.6	6.2
Monthly household income			
< 100,000 pesos	35.0	14.7	0.1
100,000–200,000 pesos	41.1	32.0	0.7
200,001–300,000 pesos	12.5	18.7	1.4
301,000–400,000 pesos	5.1	11.1	2.0
> 400,000 pesos	6.3	24.6	95.7
Monthly school fee			
None	71.4	22.5	0.6
1–4,999 pesos	17.1	10.1	0
5,000–10,000 pesos	9.8	22.8	0.2
10,001–20,000 pesos	1.4	23.8	0.2
> 20,000 pesos	0.3	20.8	99.0
Parent-reported requirements for school enrollment			
Civil marriage certificate	4.8	8.8	23.1
Church marriage certificate or certificate of baptism	1.2	17.4	34.7
Child attended play session	1.8	5.7	27.6
Child took entrance exam	7.6	43.4	62.7
Parent interview	12.1	33.8	74.7

Source: 2005 SIMCE test.
a. The 2005 survey tested 259,852 fourth graders and included 233,458 parent responses for income, 232,667 for school fees, and 240,554 for enrollment requirements.

School Choice and Its Impact on Stratification and Achievement

Economists expect that school choice aligns the incentives of school administrators and teachers with parental demand, thus improving productivity.[11] Recent theories of school markets suggest that choice might also increase stratification by income or other student characteristics. For example, Epple and Romano, on the one hand, and Urquiola and Verhoogen, on the other, present models which, despite emphasizing rather different mechanisms, hold that a system like Chile's would tend to produce sorting in terms of income or ability.[12]

11. Friedman (1955); Chubb and Moe (1990); Hoxby (2000).
12. Epple and Romano (1998, 2002) emphasize peer group effects in a setting in which students differ on both ability and income, while Urquiola and Verhoogen (forthcoming) focus on heterogeneous preferences over class size when students differ only on income.

In this section, we first review the main findings from previous research regarding the effects of choice on education outcomes and stratification in Chile. Second, we present new evidence using an alternate empirical approach focusing on census districts in Chile's Sixth Region.

Previous Evidence

The most common way in which researchers have tried to evaluate the effects of school choice on student learning is by comparing the achievement of students who attend public and private schools. In some countries, though not in Chile, analysts have been able to address this question with the benefit of randomized designs.[13] While the small-scale experiments on which such research relies are well suited to analyzing partial equilibrium questions (such as what would happen to one student's outcomes if she shifted from the public to the private sector), such experiments are not suited to answering general equilibrium questions of the kind raised by Chile's experience (such as what would happen if many students shifted from the public to the private sector).[14] In the absence of true experiments, researchers have adopted quasi-experimental approaches to identify the general equilibrium effects of choice.[15]

In one of the earliest attempts, Hsieh and Urquiola apply a difference-in-differences approach to municipal-level data from 1982 to 1996.[16] The study suggests that municipalities that experienced faster growth in private school market shares show distinct signs of stratification consistent with so-called cream skimming (that is, the middle class exiting public schools) and do not

13. For example, Angrist and others (2002) find beneficial effects of private school scholarships in Colombia. In the United States, the evidence is significantly more mixed. Rouse (1998) finds evidence of rather modest effects in Milwaukee. See also the debate between Howell and Peterson (2002) and Krueger and Zhu (2004). For examples of nonexperimental, private/public comparisons of achievement in Chile, see Aedo and Larrañaga (1994); Bravo, Contreras, and Sanhueza (1999); McEwan and Carnoy (2000); Mizala and Romaguera (2000); McEwan (2001); Contreras (2002); Anand, Mizala, and Repetto (2006); Bellei (2007).

14. McEwan (2000b); Hsieh and Urquiola (2003). More specifically, if some input is behind the private advantage, but is also in fixed supply, then these two analyses can suggest rather different impacts of choice. For example, suppose that peer quality is better in the private sector, and that this generates a causal private advantage. As the private sector expands (as happened in Chile), its peer quality advantage will dissipate, and this could lower average achievement depending on the functional form of peer effects (Hsieh and Urquiola, 2003).

15. For studies that explore the effects of choice on average achievement and stratification in the United States, see Hoxby (2000); Urquiola (2005); Rothstein (2006).

16. Hsieh and Urquiola (2003, 2006).

seem to have experienced higher growth in average outcomes.[17] The key weakness with these estimates—despite the use of some candidate instrumental variables—is that private entry into school markets is endogenous. For instance, if outcomes had been declining in areas where the private sector grew more, these effects would underestimate the salutary effects of competition.

Auguste and Valenzuela analyze the 2000 cross-section of SIMCE data, using distance to a nearby city as an instrument for the private market share.[18] They also find evidence of sorting or cream skimming, but in contrast to Hsieh and Urquiola, they find significant positive effects on achievement. The key assumption in their analysis—that distance affects education outcomes only through its effect on private school entry—is not beyond suspicion, however. It is possible, for example, that more motivated parents migrate toward cities in search of better schools.

Gallego focuses on the 2002 cross-section of SIMCE data, using the density of priests per diocese as an instrument for the number of subsidized schools relative to public schools.[19] He finds substantial effects of the competition proxy on average student achievement. An issue to consider is that priests are unlikely to have ever been randomly allocated. For instance, to the extent that they tend to be concentrated in urban areas, concerns emerge similar to those in Auguste and Valenzuela.[20] Unlike the two previous studies, Gallego does not analyze the effects of private entry on stratification.

All of the above results are reduced form, meaning they are, at best, informative of the net effect of choice on achievement and stratification, and they do not reveal the channels through which these effects might have operated. For instance, stratification could stem from certain schools' actively seeking to screen out poor or low-performing children or from the parents' of such children choosing not to apply to these institutions. The distinction between

17. McEwan (2000a) reports difference-in-differences results on achievement that are consistent with these findings, using a pooled sample of public and private school observations from 1982 to 1996.

18. Auguste and Valenzuela (2006).

19. Gallego (2006).

20. A further issue is that Gallego's instrument, unlike the one used by Auguste and Valenzuela (2006), does not correspond neatly with local school markets. There are currently twenty-six dioceses in Chile (versus more than 300 municipalities, for example). A given market (for example, the capital of a municipality) could have a relatively high number of priests despite being in a low-density diocese.

these processes is subtle and very difficult to identify, and the above strategies provide almost no leverage in this regard.

In short, the literature assessing the effect of school choice on average outcomes in Chile continues to be controversial (less so in the case of the effects on stratification). However, this evidence needs to be considered alongside aggregate levels and trends. If the effects of choice on average school productivity were indeed large, then one would expect Chile's relative performance on international tests to have improved over time.[21] Furthermore, one would expect Chile to outperform other countries with similar levels of gross domestic product per capita. None of these predictions are borne out by empirical research.[22]

New Evidence from Census Districts

Prior research is hard-pressed to identify exogenous variation in the entry of private schools. To cast further light on these issues, we rely on the discontinuous relationship between the size of a local schooling market—proxied by the population of school-aged children—and the probability that a private school enters that market. This approach is motivated by the observation that private schools are more likely to locate in heavily populated areas.[23] By itself, this provides little leverage to identify the causal impact of private school entry, since population is also correlated with variables unobserved by the researcher. However, theory from industrial organization implies a discontinuous relationship between population and private school entry, as private schools pass a profitability threshold.[24] The intuition is that a private school will only enter a market in which a public school already exists if the market is large enough to sustain two schools whose production functions display economies of scale (that is, firms need to be large enough to break even).

Suppose that private schools only enter local markets when the market size exceeds a population threshold. One might assume that local schooling markets—actually small towns—are similar in small intervals on either side of this threshold, both along variables observed by the researcher (such as household incomes) and along those that might be unobserved. To estimate

21. As this paper went to press, results from PISA 2006 were publicized, suggesting Chile might finally be on a trend of relative improvement. We do not analyze those results here owing to timing issues; furthermore, whether this trend is substantial and sustained remains to be seen.

22. Hsieh and Urquiola (2003).

23. Auguste and Valenzuela (2006); Hsieh and Urquiola (2006).

24. Bresnahan and Reiss (1987, 1991).

the impact of private school entry on stratification or achievement, one could therefore compare the outcomes of treated local markets just above the threshold to their untreated counterparts just below.[25] This is a naturally occurring variant of the regression-discontinuity design in which an assignment variable (local market size) and assignment cutoffs (population thresholds) determine whether local markets are treated by private school entry.[26]

The implementation of this approach hinges on identifying local schooling markets. Previous research on Chile almost always aggregates data at the municipal level. These jurisdictions can be large, and they often encompass many dispersed communities that are in fact separate school markets, especially in rural areas. Thus, they are not an adequate unit with which to capture the type of isolated communities that will contain the potentially discontinuous variation required by the research design.[27]

To address this challenge, we collected pilot data on the precise location of primary schools in Chile's Sixth Region, which is located south of Santiago (its geographical extension is 3,200 square miles, about twice the size of Rhode Island).[28] We merged schools' locations with geocoded data from Chile's 2002 population census, including the boundaries of 187 local census districts that fall within thirty-three larger municipalities. Table 2 reports descriptive statistics on the sample. The median census district contains two schools, but 34 percent of them contain only one. Only one district has a private enrollment rate above 60 percent.

Panel A of figure 1 presents two variables on separate *y* axes: household income and average fourth grade (math and language) test scores, both graphed against fourth grade enrollment. In Chile, enrollment in primary grades is close to universal.[29] We may therefore interpret the districtwide fourth grade

25. In practical terms, many authors estimate the average difference in outcomes, local to the threshold, by regressing the dependent variable on smooth polynomials of the assignment variable and a dummy variable indicating values above the threshold (van der Klaauw 2002).

26. The regression-discontinuity design has been applied to a range of issues in the economics of education. In Chile, it has been used to study the impact of the P-900 program, discussed below (Chay, McEwan, and Urquiola 2005), and the effect of delayed primary school enrollment (McEwan and Shapiro 2008).

27. An alternative is to use smaller jurisdictions, such as the more compact *localidad* identified in some Chilean data sets. However, the *localidad* is an informal unit and is thus inconsistently applied by the National Institute of Statistics (INE) and the Ministry of Education.

28. The geographic data collection was funded by Columbia University's Institute for Social and Economic Research and Policy (ISERP) and implemented by a private firm. We also drew on existing geographic data collected by Chile's National Institute of Statistics (INE) and the government of the Sixth Region.

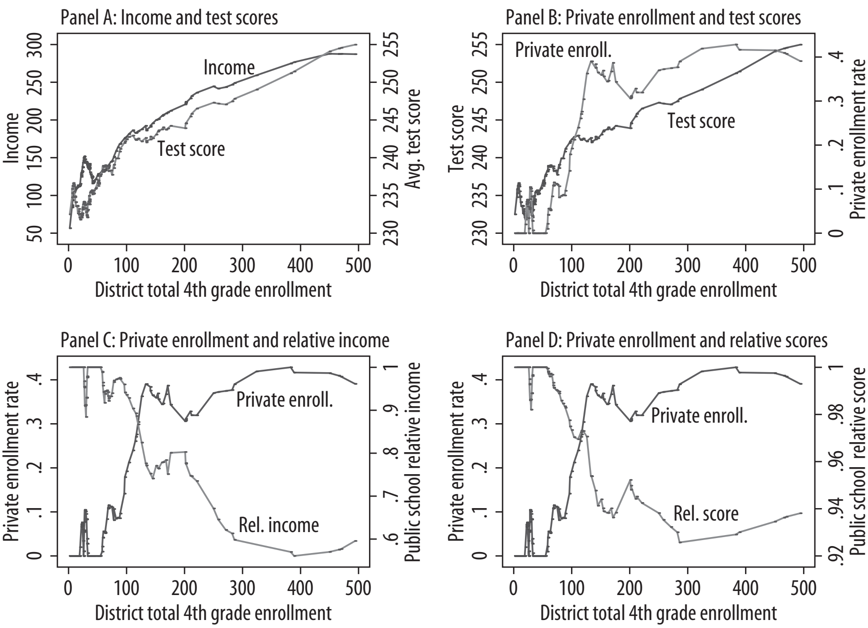
29. Urquiola and Calderón (2006).

TABLE 2. Descriptive Statistics, Sixth Region Census District Sample

Variable	*Mean*	*Std. dev.*	*Minimum*	*Maximum*	*No. observations*
Total fourth grade enrollment	80.7	116.1	2.0	708.0	187
Fourth grade enroll. in public schools	57.7	70.1	0.0	421.0	187
Fourth grade enroll. in subsidized private schools	19.6	56.4	0.0	478.0	187
Fourth grade enroll. in unsubsidized private schools	3.5	18.0	0.0	161.0	187
Number of public schools	2.2	1.5	0.0	10.0	187
Number of subsidized private schools	0.5	1.2	0.0	11.0	187
Number of unsubsidized private schools	0.1	0.4	0.0	3.0	187
Private enrollment rate	0.103	0.148	0	0.577	187
Income	153.4	116.6	50.0	888.0	172
Average test score (math and language)	238.1	21.4	192.9	311.6	172
Relative income (public to total)	0.92	0.19	0.14	1.00	168
Relative mothers' schooling (public to total)	0.97	0.07	0.61	1.00	168
Relative test score (public to total)	0.98	0.04	0.78	1.00	168

Source: Authors' calculations, using geocoded school location, census, and SIMCE data for the Sixth Region.

FIGURE 1. Market Size, Private Enrollment, Income, and Test Scores



Source: 2002 Census, 2002 SIMCE test, enrollment files, and authors' calculations.

enrollments as a proxy for market size. Both variables rise steadily with district enrollments, suggesting higher average incomes and test scores in larger communities, as observed in many other countries.[30] Furthermore, the overlap between these two lines confirms a correlation between income and test scores.

Panel B copies the segment related to test scores from panel A and adds a segment describing the relation between the private enrollment rate of census districts and our measure of market size. The segment shows that the private share is close to zero in the districts with the fewest students, and it rises sharply in districts with roughly 100 to 150 students. While the change is not altogether discrete, as would be ideal, the share reaches almost 40 percent at around 150 students and remains stable thereafter. The panel provides no evidence of a sharp change in test scores between 100 and 150 students, the vicinity of the relatively sharp rise in private enrollment. If the entry of a more effective private school (and any associated public response) caused school productivity to rise substantially in this range, one might expect test scores and income to diverge in the 100–150 student range. In fact, panel A reveals that they track each other relatively closely in this range.

Panels C and D present evidence that private school entry is associated with discrete changes in the amount of within-market stratification by family income and test scores. In panel C, the *y* axis variables are the private enrollment rate and the ratio of average public student income to the average among all students (so the ratio equals one in census districts with no private schools).[31] Panel C illustrates that this ratio declines sharply in precisely the enrollment range at which private enrollment jumps, suggesting that relatively higher-income students depart public schools once they have a private alternative. In panel D, the same pattern is evident for relative public school test scores.

The evidence in figure 1 suggests that private school entry increases within-district stratification, with no corresponding (sharp) change in average test scores. This finding is consistent with previous evidence from Chile indicating that school choice led to sorting, with no clear improvements in average achievement. We only view the current evidence as suggestive, however, given two caveats. First, it is drawn from a small sample in Chile's Sixth Region, and it remains to be seen whether the patterns would be robust to the inclusion of additional census districts and the use of regression analyses

30. See, for instance, the correlations between market size and socioeconomic status indicated by Angrist and Lavy (1999) and Urquiola (2006) for Israel and Bolivia, respectively.

31. For similar analyses at the municipal level, see Hsieh and Urquiola (2003, 2006).

(which are difficult to conduct with our reduced sample sizes). Second, the evidence is best generalized to small rural communities, although the pattern of results is consistent with evidence from a broader sample and different methods.[32]

Using Test Scores to Improve Schools: Cautionary Tales

Chile has emerged as a regional leader in the application of standardized achievement tests and the dissemination of their results. For instance, schools' unadjusted average SIMCE scores have been published in Chilean newspapers since the mid-1990s. Economists envision three roles for such information, particularly in a liberalized schooling market. First, test score data may assist parents in choosing effective (or high value added) schools. Second, and more recently, governments have used functions of schools' measured performance to directly allocate rewards, sanctions, and assistance. The most prominent example is the federal No Child Left Behind Act in the United States.[33] In Chile, the Ministry of Education has also used test scores to allocate aid and rewards since the 1990s. For example, schools with low average test scores received additional resources and training in the P-900 program, implemented in the 1990s, while schools with high scores receive merit bonuses under the ongoing *Sistema Nacional de Evaluación de Desempeño de Establecimientos Educacionales* (SNED).[34] Third, economists increasingly use student assessment data to evaluate the impact of education programs and policies. The following subsections describe three reasons why some caution is warranted in using test score data to support each role.

Test Score Levels Reflect Socioeconomic Status

One of the simplest measures of school performance is the unadjusted cross-sectional mean (or level) of students' test scores. This measure undoubtedly contains information on the contribution of schools, but as the Coleman Report and related studies revealed, it also reflects the influence of students' socioeconomic status.[35]

32. Hsieh and Urquiola (2006).
33. Hanushek and Raymond (2005).
34. On the P-900 program, see Chay, McEwan, and Urquiola (2005); on SNED, see Mizala and Urquiola (2008).
35. Coleman and others (1966).

Consequently, in a context of extensive school-based stratification, school rankings based on unadjusted average test score levels might mainly reflect student socioeconomic status. Mizala, Romaguera, and Urquiola examine this issue with eight rounds of SIMCE data between 1997 and 2004, using a sample of 701 schools for which test score and socioeconomic data are consistently available.[36] In each case, schools' average language scores are regressed on mothers' and fathers' schooling and household income.[37] Depending on the year, between 70 and 85 percent of the variation in test scores can be explained by this small set of observable characteristics, suggesting that a ranking based on schools' average test scores may not be very different from one based on students' socioeconomic information. To explore this possibility, the authors consider two hypothetical programs that selected the top fifth of performers using different rules: one ranks schools according to their mean language score in SIMCE data, and the other ranks schools according to average mothers' schooling, reported in the same surveys. They show that these rules agree on the selection or exclusion of about 85 percent of all schools; a similar correspondence results for a program that selected the bottom fifth of performers.

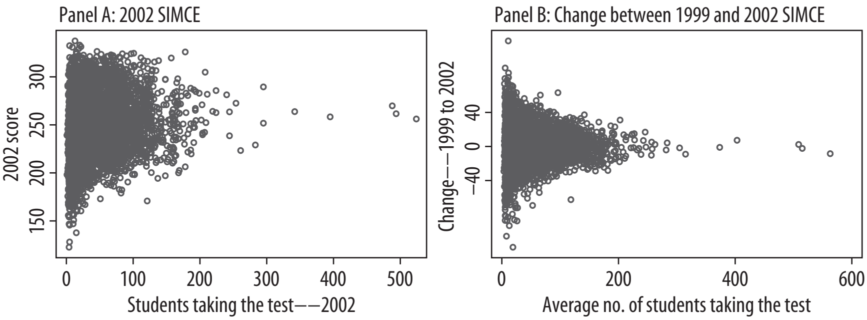
These findings raise two issues. First, it is not clear that schools' test score levels are correlated with their true value added.[38] For example, one might imagine an exceptionally productive school that enrolls predominantly low-income children or, likewise, a rich school with unexceptional administrators and teachers. In such cases, access to information on test score levels may not improve parents' ability to choose the most productive schools. Indeed, it may simply reinforce stratification or reward inefficiency if it encourages parents to choose low value added schools serving wealthy children. This could be entirely consistent with rational parents' desire to improve their children's achievement, however, if they believe that exposure to wealthy peers could improve longer-run outcomes. Second, government accountability programs that purport to identify, reward, or sanction "good" or "bad" schools—that is, those with high or low value added—can misallocate resources if they rely exclusively on test score levels.

36. Mizala, Romaguera, and Urquiola (2007).

37. For complete details on variables and specifications, see Mizala, Romaguera, and Urquiola (2007).

38. Below, we show that schools with lower test score levels in 1988 actually have higher gains in subsequent years (where gain scores are one proxy for schools' value added). However, we suggest that this is an artifact of testing noise and mean reversion, rather than schools' true value added.

FIGURE 2. School Performance and School Size: Fourth Grade Language Scores



Source: 1999 and 2002 SIMCE language scores and authors' calculations.

Controlling for Socioeconomic Status Can Raise Volatility of School Rankings

A reasonable response to the problems of using unadjusted test scores is to use additional data and methods to control for students' socioeconomic status and obtain a better estimate of schools' value added.[39] One challenging aspect of such analyses is that many family and student variables are unobserved by the researcher, and their omission from statistical specifications potentially biases estimates of school effects. A less well known but nonetheless important challenge, however, is that controlling for socioeconomic status often comes at the expense of introducing volatility into the test score rankings of schools, especially in a stratified school system like Chile's.[40]

What might produce volatility in schools' test performance? First, one-time events, such as a schoolwide illness or distraction on the testing day, could influence test scores.[41] Second, there is sampling variation, since each cohort of students that enters a school is analogous to a random draw from a local population. A school's average performance will thus vary with the specific group of students starting school in any given year. This variance, in turn, depends on the variability of performance in the population from which the school draws its students and the number of students tested.

Chay, McEwan, and Urquiola verify the implication that scores should be more variable in schools with lower enrollment.[42] Figure 2 (panel A) pre-

39. Hanushek and Raymond (2003).
40. Mizala, Romaguera, and Urquiola (2007).
41. Kane and Staiger (2001).
42. Chay, McEwan, and Urquiola (2005).

sents a similar analysis using fourth grade test scores from the 2002 SIMCE. Smaller schools tend to display greater variability in their performance, where school enrollment (on the *x* axis) is measured by the number of fourth grade students who took the test.

Test-score volatility is even more pronounced when we use measures that arguably do a better job of controlling for student socioeconomic status, such as changes in school performance between separate cohorts of fourth graders (1999 and 2002). This measure is appealing to the extent that it controls for time-invariant characteristics of schools, and it better approximates value added.[43] The measure is also relevant since it is used in accountability schemes throughout the world.[44] Panel B in figure 2 shows that the inverse relationship between school size and performance variability becomes even more pronounced—smaller schools on average improved no more or less than larger schools, but their performance is much more volatile.

To explore how volatility in school test scores affects school rankings, we consider a fictitious program that identifies the top 20 percent of schools in each of eight years (see table 3).[45] As a benchmark, suppose that schools' performance does not vary over time. The column labeled "certainty" contains the distribution that one would observe if schools' performance were stable: 80 percent of schools would never appear in the top quintile, and 20 percent would be in this category all eight years. The opposite benchmark supplies the percentages that would be observed if the program were a lottery—that is, if all schools had independent 0.2 probabilities of being selected each year. In this case, one would expect that over the eight years, 17 percent of schools would never get selected. The modal school would end up in the top quintile only once, and no school would be selected every year. The table also summarizes the results that emerge using actual test score levels over eight years. The observed distribution is fairly close to the distribution under stable school performance. This is not surprising given that, as stated, this measure substantially reflects socioeconomic status, and that schools' position in the relative socioeconomic status distribution is relatively stable over time. Finally, the table presents the results of using the eight cross-sections of data to calculate seven observations of changes in mean test scores. The actual distribution in this case is very close to that of a lottery. Over eight years, more than

43. Mizala, Romaguera, and Urquiola (2007) also explore the use of regression residuals left after controlling for observed socioeconomic status.

44. Hanushek and Raymond (2003).

45. This exercise follows Kane and Staiger (2001).

TABLE 3. Frequencies in Top or Bottom 20 Percent Produced by Different Rankings

	Levels			*Change in test score*		
Frequency	*Certainty*	*Lottery*	*Actual*	*Certainty*	*Lottery*	*Actual*
A. No. times school in top 20 percent						
Never	80.0	16.7	59.3	80.0	21.0	16.6
1 year	0.0	33.6	9.1	0.0	36.7	40.0
2 years	0.0	29.4	5.0	0.0	27.5	31.7
3 years	0.0	14.7	5.7	0.0	11.5	10.7
4 years	0.0	4.6	5.3	0.0	2.9	1.4
5 years	0.0	0.9	3.0	0.0	0.4	0.0
6 years	0.0	0.0	3.9	0.0	0.0	0.0
7 years	0.0	0.0	4.1	20.0	0.0	0.0
All 8 years	20.0	0.0	4.6	. . .	. . .	. . .
B. No. times school in bottom 20 percent						
Never	80.0	16.7	68.1	80.0	21.0	19.1
1 year	0.0	33.6	5.1	0.0	36.7	34.7
2 years	0.0	29.4	2.3	0.0	27.5	33.5
3 years	0.0	14.7	1.4	0.0	11.5	11.7
4 years	0.0	4.6	3.9	0.0	2.9	1.0
5 years	0.0	0.9	3.9	0.0	0.4	0.0
6 years	0.0	0.0	2.6	0.0	0.0	0.0
7 years	0.0	0.0	5.9	20.0	0.0	0.0
All 8 years	20.0	0.0	7.0	. . .	. . .	. . .
C. Schools ever in top and bottom 20 percent						
Percentage	. . .	. . .	0.3	. . .	. . .	72.5

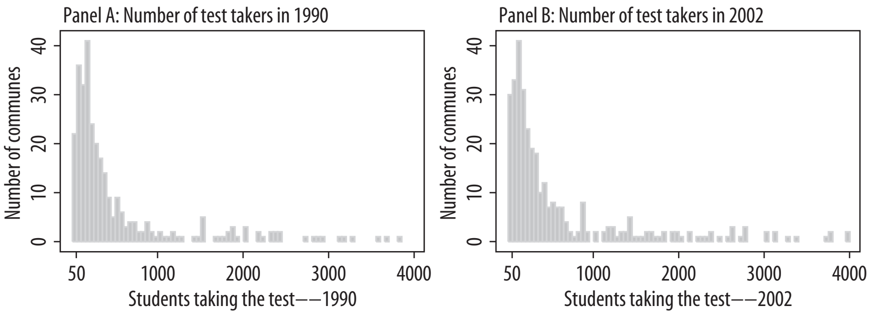
Source: Mizala, Romaguera, and Urquiola (2007).
. . . Not applicable.

80 percent of schools would have been in the top performing group at some point, and more than 70 percent would have been in both the best and worst performing groups.

This last finding suggests that employing measures that better control for socioeconomic status results in an undesirable trade-off. While the school rankings may contain some signal of school effectiveness, the added volatility raises concerns. First, ever-shifting conclusions about the "best" or "worst" schools may lead parents to sort across schools inefficiently. Hanushek, Kain, and Rivkin, for instance, argue that higher student turnover can entail significant negative externalities.[46] Second, teachers and school administrators might begin to respond less to incentive programs whose ex post allocations resemble a lottery.

46. Hanushek, Kain, and Rivkin (2004).

FIGURE 3. Municipality Size Measured by the Number of Fourth Grade Test Takers, 1990 and 2002[a]



Source: 1990 and 2002 SIMCE test and authors' calculations.

a. For clarity, the figures exclude between five and ten municipalities in which more than 4,000 students took the respective exam.

Volatility Concerns Persist with Aggregate Data

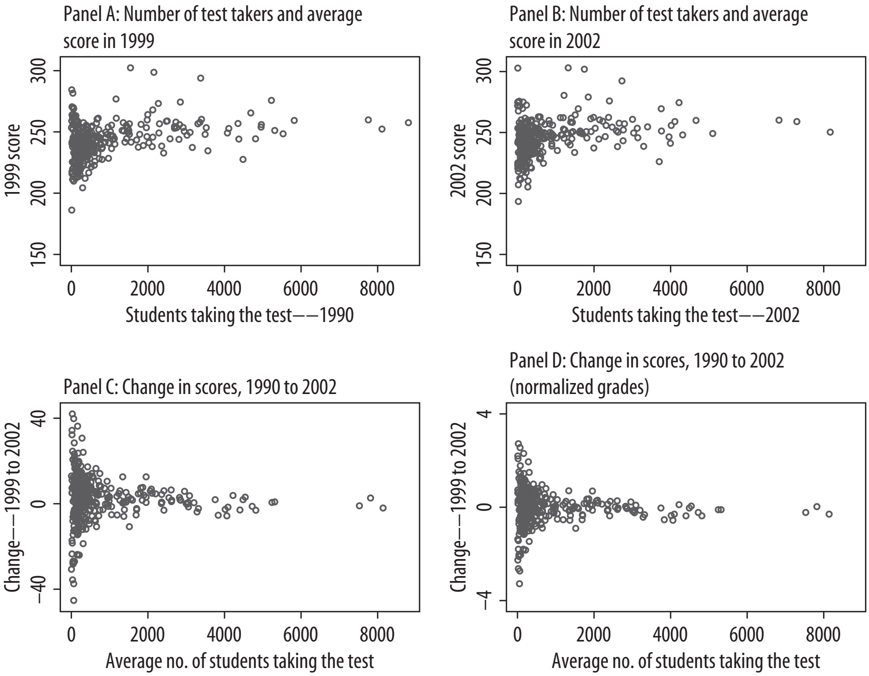
The concerns highlighted above have been raised mostly in the context of school- or classroom-level information. There is consensus that these issues would be mitigated if data were aggregated to higher levels, such as cities or states. The intuition is that averaging across large numbers of schools or students will reduce the influence of one-time, school-specific events, as well as of sampling variation among smaller schools.

In Chile, the municipality is a policy-relevant unit of aggregation at least for public schools, since municipalities are responsible for school management. For this reason, Engel, Mizala, and Romaguera list the most and least improved municipalities in terms of SIMCE performance, and they call on voters to hold their mayors accountable for this performance.[47]

One must single out such "good" or "bad" municipalities with care, however, because volatility concerns persist even with data aggregated to this level. Many municipalities are small as measured by the number of children that take the SIMCE test in a given year. Panel A in figure 3 shows that in 1990, fifty or fewer students took the fourth grade SIMCE test in twenty-one of 315 municipalities, and a hundred or fewer students took it in fifty-four municipalities. While there are also large municipalities (sixty-six had more

47. Eduardo Engel, Alejandra Mizala, and Pilar Romaguera, "Educación municipal: qué tal lo ha hecho su alcalde?" *La Tercera,* 19 September 2004.

FIGURE 4. Average Language Score and Municipality Size, 1999 and 2002



Source: SIMCE, 1990 and 2002, and authors' calculations.

than 1,000 test takers), the data indicate that several municipalities have fewer students than a large school in Santiago.[48] Panel B in figure 3 presents similar data for 2002. Despite a decade's worth of population growth, twenty-six of 337 municipalities had fewer than fifty test takers, and sixty-two had fewer than a hundred.

Figure 4 shows that there is more variation in the performance of smaller jurisdictions than larger ones at the municipal level. Panel A presents the average language score in each municipality in 1999 ordered against the total number of test takers; panel B presents similar evidence for 2002. In both years, the general inverse relationship between size and performance variability is observed. Panel C focuses on the change in scores between 1999 and

48. The mean municipality has approximately 730 students, with a standard deviation of about 1,150.

2002 and shows that, as with school-level data, extreme cases of municipal-level improvement or worsening are most likely found among the smaller jurisdictions.

Table 4 shows that small municipalities dominate in a listing of the ten best and worst municipalities in language performance in 1990 and 2002. In both years, the list of top performers selected from the full sample features municipalities with high levels of income and parental schooling, such as some of the wealthier neighborhoods in Santiago (namely, Providencia, Vitacura, and Las Condes). Some extremely small municipalities like Torres del Paine (in 1990) and Primavera are also among the top performers—with five and seventeen test takers, respectively. Very small municipalities also figure among the worst performers. For comparison, table 4 includes lists that restrict the sample to ninety-seven municipalities with at least 350 students taking the SIMCE in both 1990 and 2002.

Finally, table 5, analogous to the school-level table 3, uses SIMCE test scores for 1990, 1992, 1994, 1996, 1999, and 2002 (that is, the years since 1990 in which the SIMCE was administered at the fourth grade level) to explore the extent to which issues of ranking volatility persist with municipal-level data. The results show that while rankings based on municipalities' absolute scores are relatively stable (even more so than with schools), rankings using changes in scores are still hard to distinguish from a lottery.

Volatile School Rankings Can Bias Evaluations

Economists are increasingly interested in evaluating school interventions that are assigned on the basis of test scores, such as the rewards and sanctions distributed by school accountability programs. The volatility of test-based measures raised in the previous section can affect these evaluations. Consider a ranking of schools by their mean test scores, used to identify the "worst" schools for inclusion in a program. Our discussion suggests that such schools are likely to be disproportionately small, as a result of sampling variation and an unlucky draw of students. Unless the bad luck is systematically repeated in the next year—which is unlikely if random events or sampling variation are the culprit—then we would expect the average score among these schools to bounce back in subsequent years. This mean reversion can easily be mistaken for a positive program effect. In the opposite case, schools with extremely positive scores in one year might dip in the following year, which could be mistaken for a negative program effect. In both cases, common evaluation approaches can over- or underestimate the size of program effects.

TABLE 4. The Best- and Worst-Performing Municipalities in 1990 and 2002

	Full sample			*Large municipalities*			
Year and ranking	*Municipality*	*Grade*	*No. students*	*Ranking*	*Municipality*	*Grade*	*No. students*
1990 results							
1	Providencia	80.5	2,197	1	Providencia	80.5	2,197
2	Torres del Paine	79.3	5	2	Las Condes	76.4	4,412
3	Las Condes	76.4	4,412	3	Ñuñoa	71.7	2,915
4	Pumanque	74.6	65	4	La Reina	71.6	1,522
5	Primavera	74.0	17	5	Santiago	68.8	7,479
6	Ñuñoa	71.7	2,915	6	Punta Arenas	67.0	1,910
7	La Reina	71.6	1,522	7	Arica	66.9	3,270
8	Navarino	71.1	36	8	Villa Alemana	66.5	1,058
9	Porvenir	70.1	87	9	Coyhaique	65.8	876
10	Santiago	68.8	7,479	10	Iquique	65.7	2,886
. . .				. . .			
306	Lumaco	42.9	166	88	Monte Patria	51.7	504
307	Ercilla	42.7	173	89	Calbuco	51.6	368
308	Chaco	42.3	137	90	Paine	51.1	624
309	Lonquimay	41.2	172	91	Nueva Imperial	50.5	390
310	Quellen	40.7	45	92	La Pintana	50.5	2,343
311	Curarrehue	40.3	77	93	Cañete	50.2	674
312	Saavedra	40.1	139	94	Collipulli	49.7	438
313	Tirua	39.1	246	95	Panguipulli	48.9	603
314	Empedrado	37.4	94	96	Los Alamos	48.6	452
315	Colchane	35.1	12	97	Longavi	46.5	493
2002 results							
1	Vitacura	303.2	1,330	1	Vitacura	303.0	1,330
2	Providencia	299.1	1,758	2	Providencia	299.1	1,758
3	Primavera	297.3	10	3	Las Condes	291.9	2,733
4	Las Condes	291.9	2,733	4	Lo Barnechea	279.0	1,211
5	Lo Barnechea	279.4	1,211	5	La Reina	276.3	1,852
6	La Reina	276.3	1852	6	Ñuñoa	274.1	2,398
7	Navarino	276.1	45	7	Santiago	272.6	4,215
8	Ñuñoa	274.1	2,398	8	San Miguel	267.3	1,585
9	Santiago	272.6	4,215	9	Concepcion	267.1	3,756
10	Hualañe	271.6	149	10	Macul	262.0	1,505
. . .				. . .			
324	Yerbas Buenas	216.7	172	88	Graneros	233.7	486
325	Quilaco	215.7	53.5	89	Lo Espejo	232.5	1,503
326	S. Juan de la Costa	215.4	143	90	Victoria	231.6	664
327	Tolten	212.7	181	91	Padre las Casas	231.4	1,046
328	Galvarino	211.5	224	92	Monte Patria	230.9	531
329	La Higuera	210.1	59	93	Longavi	230.0	416
330	Huara	209.4	28	94	Lautaro	229.0	586
331	Til-Til	209.3	279	95	Panguipulli	226.0	596
332	Saavedra	201.9	272	96	La Pintana	225.6	3,691
333	Colchane	191.4	23	97	Nueva Imperial	224.0	578

Source: 1990 and 2002 SIMCE test and authors' calculations.

TABLE 5. Frequencies in Top or Bottom 20 Percent Produced by Different Rankings, Municipal-Level Observations[a]

Frequency	*Levels*			*Change in relative score*		
	Certainty	*Lottery*	*Actual*	*Certainty*	*Lottery*	*Actual*
A. No. times school in top 20 percent						
Never	80	26.2	56.7	80	32.8	32.5
1 year	0	39.3	11.4	0	41.0	34.2
2 years	0	24.6	7.3	0	20.5	20.8
3 years	0	8.2	5.6	0	5.1	6.4
4 years	0	1.5	5.6	0	1.0	2.9
5 years	0	0.1	4.7	20	0.0	3.2
All 6 years	20	0.0	8.8	. . .	. . .	. . .
B. No. times school in bottom 20 percent						
Never	80	26.2	57.6	80	32.8	36.6
1 year	0	39.3	14.9	0	41.0	36.6
2 years	0	24.6	8.5	0	20.5	23.7
3 years	0	8.2	6.1	0	5.1	3.2
4 years	0	1.5	3.8	0	1.0	0.0
5 years	0	0.1	6.1	20	0.0	0.0
All 6 years	20	0.0	2.9	. . .	. . .	. . .

Source: SIMCE tests (1990, 1992, 1994, 1996, 1999, and 2002) and authors' calculations.

. . . Not applicable.

a. Panel A refers to a hypothetical program that would select the top 20 percent of municipalities, and panel B presents an analogous exercise for a program that would select the bottom 20 percent. Columns 1 and 4 describe the distribution that would be observed over the six years under "certainty," for test score levels and changes, respectively. Columns 2 and 5 provide the distribution that would be obtained if the program were a lottery, for levels and changes, respectively.

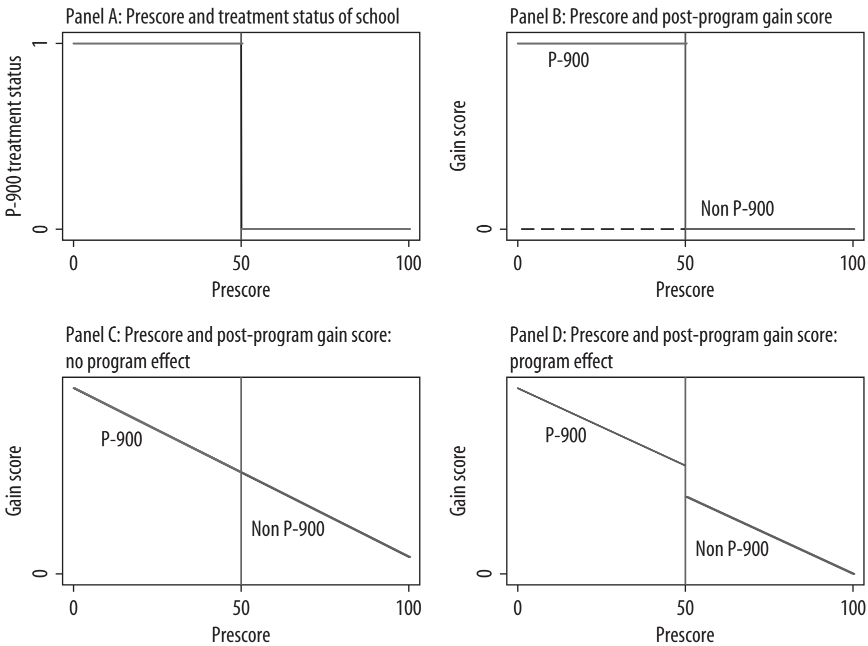
In the early 1990s, Chile's 900 Schools Program (or P-900), later renamed and modified by the Ministry of Education, consisted of a targeted intervention to a group of schools that received improvements in infrastructure, instructional materials, teacher training workshops, and after-school tutoring workshops for children not performing at grade level.[49] The program's initial assignment, in the 1990 school year, was primarily based on 1988 SIMCE test scores. Schools in thirteen regions were ranked from best to worst. Separate cutoff scores were established for each region, and schools that scored below the cutoff score were preselected to participate, although some were subsequently removed.[50]

Panel A in figure 5 illustrates a stylized version of the assignment rule by plotting, on the *x* axis, the average score of each school in the year used to determine school rankings (called the prescore) and, on the *y* axis, the treatment status of the school, which takes a value of 0 or 1. The prescores range

49. Chay, McEwan, and Urquiola (2005).

50. This removal has important methodological implications that we gloss over here; they are discussed at length in Chay, McEwan, and Urquiola (2005).

FIGURE 5. Hypothetical Assignment of P-900 Schools and Effects on Test Scores



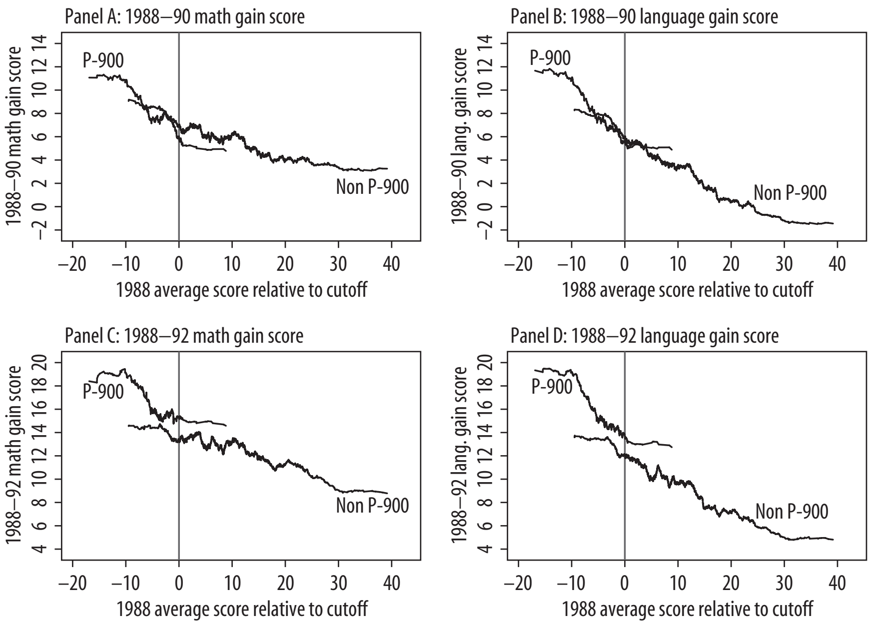
Source: Chay, McEwan, and Urquiola (2005).

from 0 to 100, and we arbitrarily choose 50 as the program cutoff, such that all schools with prescores of 50 or less are treated with the P-900 program, and the rest are not.

P-900 has been praised as one of the most successful school interventions in the developing world, and the literature reflects a perception that it substantially raised the achievement of treated schools.[51] Specifically, previous research finds that treated schools' average test scores improved more than those of untreated schools, a result obtained from a difference-in-differences specification. For example, the 1988 to 1990 changes in scores imply program effects that are equivalent to 0.3 and 0.5 standard deviations in math and language, respectively. For 1988 to 1992 gain scores, the implied P-900 effects are 0.5 and 0.7 standard deviations, respectively. For these estimates to have a causal interpretation, the differences in the gain scores of treated and untreated schools must be entirely due to the program. Panel B of fig-

51. See, for example, García-Huidobro (1994, 2000).

FIGURE 6. Gain Scores by 1988 Average Score Relative to the Regional Cutoff



Source: Chay, McEwan, and Urquiola (2005).

ure 5 illustrates a scenario in which this may be the case. The *y* and *x* axes display the post-program gain score and the prescore, respectively. Here, the vertical distance between the two line segments is the added gain among P-900 schools (that is, the treatment effect), and a causal interpretation of the effect may be justified. In contrast, if schools with lower prescores have higher gain scores even in the absence of P-900, then the program effect may be closer to zero. This is the case in panel C of figure 5, which shows no break in the relation between the gain score and the prescore close to the assignment cutoff. However, a difference-in-differences specification would erroneously identify a positive treatment effect. Finally, panel D illustrates a break in test scores close to the assignment cutoff, consistent with a program effect.

Using a regression-discontinuity approach to consider the actual evidence, we plotted unweighted smoothed values of schools' gain scores against their 1988 prescores (relative to their respective regional cutoff), distinguishing between P-900 and untreated schools; the results are presented in figure 6. We find a negative relation between gain scores and 1988 scores, which is consistent with substantial mean reversion. To the extent that P-900 had an effect, we should observe a break in these relationships close to the cutoff (one analogous to that in figure 5, panel D). Restricting the analysis to observations in this range implements a regression-discontinuity approach. The graphs for 1988–90 gain scores (figure 6, panels A and B) reveal no such break, since the P-900 and non-P-900 lines essentially overlap at the cutoff. Nevertheless, a "naïve" evaluation would suggest that P-900 had a large effect in its first year. Panels C and D, in contrast, indicate a break roughly equal to two points in the 1988–92 gain scores.[52] In the case of 1988–92 gain scores, the magnitude of P-900 effects is about one-third as large as estimates returned by difference-in-differences models.

We conclude that test score noise and the resulting mean reversion produce important complications in the evaluation of schemes using accountability-type test score rankings. The use of intuitively appealing evaluation strategies, like difference-in-differences, can lead to dramatically biased estimates of the program effects. Regression-discontinuity methods, which are often the natural by-product of accountability policies, can circumvent these biases.

Policy Discussion

The last three decades have been remarkable for education in Chile. In the 1980s, the military government introduced a large-scale school choice initiative. Since 1990, successive democratically elected center-left administrations have chosen not to alter the essential features of that scheme, but rather have focused on increasing resources and launching a series of initiatives. These include the following: targeted programs to raise quality in schools serving the poor, extending the years of mandatory basic education to twelve and ensuring that all first-grade entrants remain in schools throughout the twelve-year cycle, lengthening the school day, raising teacher pay, improving infrastructure, reforming the national curricula, establishing national standards, and introducing performance assessment systems for students, teachers, and schools.

Together, these policies have undoubtedly raised some households' welfare and improved school quality along many dimensions. Anyone visiting a

52. In fact, two types of breaks are visible in these figures: those at the cutoff and those for treated and untreated schools with overlapping 1988 scores (our initial regression evidence will capture both). Their magnitudes are generally similar. This reflects the fact that the assignment discontinuities (figure 6) are imperfect, leading to fuzziness across the cutoff score.

public or private school in Chile can see as much—relative to much of the region, classes take place reliably, many teachers seem to be satisfied with their work, and the infrastructure is adequate, even in rural areas. This reflects Chile's economic growth, but it also reflects the resources and effort expended on schools. Additionally, many households have taken advantage of choice to shift their children into subsidized private schools, and based on revealed preference, this must have improved their welfare.

Despite these advances and for reasons not yet fully understood, Chile's education system has made insufficient progress in improving student learning, as measured by test scores. Stratification probably increased over the period, and the associated perceptions of inequality were likely exacerbated when, beginning in 1994, private subsidized schools were allowed to charge supplementary tuition. Not surprisingly, schools with higher-income children tended to avail themselves of this opportunity to expand their funding.

The failure of the reforms to substantially raise learning outcomes and reduce school-based inequality has produced widespread dissatisfaction. Chile is perhaps the only country in Latin America where a majority of citizens list education quality and equity as one of their main political concerns, and education issues are regularly featured in the media. Indeed, these concerns were the main reason students called for a nationwide secondary-level student strike that took place in May and June of 2006. In 2007, a number of measures to raise the quality and equity of education were announced, including the creation of a Superintendency of Education responsible for quality assurance; a substantial increase in the value of the voucher subsidy; a change in the voucher design to begin differentiating its value according to students' socioeconomic status (tied with the introduction of further school accountability measures); and a significant overhaul of the laws that regulate the system, aimed at prohibiting publicly financed schools from selecting students, requiring more stringent reporting on the use of public funds, and prohibiting the operation of schools on a for-profit basis.

In the resulting policy debate, perhaps the single most controversial issue has been the proposal to eliminate private schools' ability to operate for profit. The existing evidence does not suggest that this policy would produce high returns in terms of education quality. Just as shifting large numbers of children into for-profit schools seems to have had a limited impact on average test scores, stripping those schools of their ability to profit is unlikely to lead to a large change. Furthermore, a key challenge in raising quality seems to be providing schools (as well as parents and students) with adequate incentives, and to the extent that for-profit schools might be responsive to well-specified incentives, it seems worth keeping them in existence. In addition, as Engel has noted, schools can devise ways to conceal profits.[53]

A second measure, changing the voucher design such that it is higher for children of lower socioeconomic status, was recently approved by Chile's Parliament. This preferential subsidy (as it is called in Chile) is intended to discourage cream skimming and increase accountability based on student learning outcomes (as opposed to just attendance).[54] It seems a promising avenue for raising quality and equity, although the literature does not provide much direct guidance on its potential impact and its details have not been fully worked out (particularly with respect to the precise gradient between socioeconomic status and the size of subsidies).

An additional proposal that would affect stratification involves prohibiting student selection by schools in grades one through eight. As stated above, selective admissions are widespread in Chile, particularly among private schools, which rely on tools ranging from testing to requiring marriage and baptismal certificates. At present, the details of this proposal are still to be worked out (for instance, whether to allow sibling preference, how the policy would address religious schools, and how compliance would be monitored), and its ultimate impact is difficult to predict given our incomplete knowledge of the channels through which stratification comes about.[55] Nevertheless, this initiative also seems worth exploring, not least because reducing stratification might improve the ability of the system to evaluate school performance. At the same time, its eventual impact on stratification is hard to gauge. For instance, some stratification happens naturally as a result of the interaction between residential sorting and the fact that children of primary age typically do not travel far to attend school. Furthermore, such segregation is potentially exacerbated in Chile by the existence of add-on tuition. Finally, much caution is needed in implementing this policy because the proposal could affect the ability of the very best public secondary schools (which rely on testing

53. Eduardo Engel, "El lucro y las políticas públicas," *La Tercera,* 15 April 2007.

54. Schools that received additional funding for lower socioeconomic status children would have to sign agreements with the Ministry of Education stipulating specific student learning goals, to which they would be held accountable.

55. Similar regulatory issues exist for publicly funded and privately managed charter schools in the United States. The majority of state-specific charter school laws explicitly require schools to admit students by lottery if their applications exceed available positions, although exceptions are often made for siblings of current students (McEwan and Olsen 2007). The conduct of many such lotteries is self-monitoring, in that they are conducted, by law or custom, in a public forum.

selection generally at the seventh grade) to excel. Some observers have noted that selective public schools have contributed to social mobility in Chile.[56]

As the Chilean government moves forward in education reform, it will need to continue to collect, disseminate, and use information on school performance. While the findings in this paper indicate reasons for caution in this regard, it is still the case that generating appropriate incentives for schools requires the use of high-quality performance data.[57] Chile would therefore do well to invest more in data collection and analysis. For instance, it could gather information on student-level gains or produce filtered estimates in the manner discussed by Kane and Staiger (although the theory and practice on how to calculate these is limited, even in the United States).[58] None of these strategies is likely to provide a silver bullet solution to the challenges described, but they might improve the quality and usefulness of school performance information.

Finally, an important issue not receiving much attention in the current debate is the need for Chile to do more to assess its own education policies and initiatives. Whenever possible, evaluations should be planned and designed in conjunction with policies, and they should employ experimental or high-quality quasi-experimental methods.[59] Over the past decades, Chilean administrations have shown imagination in designing and implementing education reform, but they have been less inclined to rigorously evaluate these initiatives. Given the universal difficulty in designing effective education policy, this must change if Chile is to optimize its ability to improve interventions and increase their impact on school quality and equity.

56. Alejandra Mizala and Pablo González, "La LOCE y la calidad de la educación," *La Tercera,* 15 April 2007.

57. Mizala and Urquiola (2008).

58. Kane and Staiger (2001).

59. Shadish, Cook, and Campbell (2002).