Distorted Quality Signals in School Markets*

José Ignacio Cuesta[†] Felipe González[‡] Cristian Larroulet[§]

Abstract

Information plays a key role in markets with consumer choice. In education, data on schools is often gathered through standardized testing. However, the use of these tests has been controversial because of distortions in the metric itself. We study the Chilean educational market and document that low-performing students are underrepresented in test days, generating distortions in school quality information. These distorted quality signals affect parents' school choice and induce misallocation of public programs. These results provide novel evidence for the costs that distortions in quality signals generated by standardized tests in accountability systems impose on educational markets.

Keywords: accountability, schools, quality, disclosure, choice
JEL Codes: I20, L15

---

*This version: April 2019. First version: March 2016. We would like to thank Claudia Allende, Gustavo Bobonis, Andreas de Barros, Michael Dinerstein, Fred Finan, Francisco Gallego, Matt Gentzkow, Michael Greenstone, Ali Hortaçsu, Brian Kisida, Edward Miguel, Magne Mogstad, Tomás Monarrez, Casey Mulligan, Derek Neal, Christopher Neilson, Nicolás Rojas, Jesse Rothstein, Jessica Scheld, Chris Walters, and seminar participants at APPAM (2017), the Bay Area Labor and Public Economics Graduate Student Conference at Stanford (2015), the Empirics and Methods Conference at Northwestern (2016), the First Conference of the RISE Programme (2015), the LACEA-LAMES Meetings (2017), the Pacific Conference for Development Economics at Stanford (2016), PUC-Chile, UC Berkeley, University of Chicago, University of Talca and the WEAI Annual Conference (2017) for comments and suggestions. We would also like to thank the Ministry of Education and the Agency for the Quality of Education of the government of Chile for access to the data and useful discussions.

†University of Chicago. Corresponding author. Email: `jicuesta@uchicago.edu`.

‡Pontificia Universidad Católica de Chile. Email: `fagonza4@uc.cl`.

§University of Colorado, Boulder. Email: `crla8025@colorado.edu`.

1 Introduction

Information plays a key role in markets with consumer choice and government intervention. In education, information on schools and students for the purposes of consumer choice and government interventions is often gathered via standardized tests, and their use has became common in recent decades (Figlio and Loeb, 2011). However, accountability systems that resort to these tests have been controversial among academics and educators. Critics argue that, because of undesirable behavioral responses and/or measurement problems that distort actual test scores, the usage of standardized tests in accountability systems is problematic (e.g. Figlio and Getzler 2002; Kane and Staiger 2002; Neal 2013). How large are distortions in test scores? And, more importantly, what are the market and policy consequences of these distortions? Despite increasing evidence of undesirable behavioral responses and measurement problems, quantification of these distortions and their consequences is surprisingly lacking.

This paper quantifies distortions in school quality signals and their market consequences. We study one of the most developed accountability systems in the world –Chile's market-oriented educational system. The Chilean government relies on standardized testing to generate school-specific quality signals. These measures are used not only for quality assessment and performance evaluation, which are key inputs in the policy process, but also as a disclosure system in school choice. These features make Chile an ideal setting to quantify the consequences of distortions in test scores on both household school choices and the allocation of public programs. We show that heterogeneous test day attendance distorts the school-specific quality metric relevant in this market. Moreover, we show that these distortions are large and have significant consequences both on school choice and on the allocation of public programs.

The analysis proceeds in four steps. First, we show that high-performing students are more likely to attend school on test days and low-performing students are more likely to stay at home. Using national administrative data on the Chilean educational system, we compare daily attendance of test takers (fourth graders) and non-takers (third graders) within schools on test and non-test days through the distribution of academic performance. High (low) performing students increase (decrease) their school attendance on test days by 3.5 (2.2) percentage points. The degree of student non-representativeness varies considerably across schools.

Second, in order to estimate the associated distortions in school-level test scores, we use a measure of academic achievement available for all students and a multiple imputation method to predict the test scores of absent students.[1] We find average distortions in the system to be sizable: 0.10 standard deviations of school test scores. Distortions vary widely across schools, but are persistent within

[1]Multiple imputation methods are routinely used in the Survey of Consumer Finances conducted by the Federal Reserve in the U.S., and in the Household Financial Survey conducted by the Central Bank of Chile, among many others (Kennickell, 1998; Alfaro and Fuenzalida, 2009).

schools over time. We provide support for our imputation approach using cross validation and accounting for selective attendance. Our analysis strongly suggests that the patterns of absenteeism on test days and consequent distortions are not random.

To better understand these distortions, we construct a panel dataset for all Chilean schools during the period 2005–2013. First, we show that fixed characteristics of schools explains a large part of distortions: public, low quality, and for-profit schools display larger average distortions. We then study the extent to which behavioral responses explain distortions in test scores. We find some evidence consistent with the hypothesis of a strategic response of schools to competitive incentives. In particular, we find that schools facing more quality-elastic consumers display larger distortions in test scores.[2] Finally, we find mixed evidence for potential perverse incentives set by public programs. On the one hand, we find that accountability pressures associated with a program that provides additional government funding to schools are associated with larger distortions. On the other hand, we find no association between distortions and public programs that provide teacher monetary incentives and that disclose school quality to households.

Third, we estimate a school choice model to quantify the implications of these distortions. We find that providing undistorted school quality information would induce three percent of students to switch schools. We develop a discrete choice model in which households trade-off school quality and distance, and estimate it using geocoded addresses of 100,000 students and 1,500 schools. For identification, we exploit quasi-experimental variation in government programs and fixed characteristics of competitors. Given the magnitude of distortions and the spatial distribution of schools, the trade-off between distance to school and quality explains the student switching rate among schools. Our results suggest that households that would change their choices are willing to pay 117 U.S. dollars annually for undistorted quality information, with high-income households willing to pay more than low-income households due to differences in preferences over school fees and school quality.

Fourth, we show that two large public programs are significantly misallocated because of distortions. In the first program, the government assigns bonuses to teachers in schools with sufficiently high average test scores. We reallocate bonuses based on removing distortions, and find that 13 percent of resources are misallocated each year, equivalent to $20 million U.S. dollars in the last twenty years. In the second program, the government used test scores to classify schools in three quality categories and delivered this information to parents with the objective of assisting school choice. Using the classification algorithm, we show that four percent of schools were incorrectly classified and these errors persuaded two percent of the incoming student cohort to choose a different school.

This paper makes three main contributions. First, we document a novel channel through which school performance measures can get distorted: relative test day attendance of high/low-performing

[2] Recent studies have suggested a link between competitive environments and cheating behavior (see Shleifer 2004; Gilpatric 2011; Cartwright and Menezes 2014, among others).

students.[3] Changes in relative test day attendance of high- and low-performing students have not been documented in the U.S., where the most common sources of distortion are manipulation of the testing pool via selective assignment of students to special education programs (Jacob, 2005; Cullen and Reback, 2006; Rockoff and Turner, 2010; Figlio and Loeb, 2011) and selective application of disciplinary measures (Figlio, 2006). Second, this is the first paper that implements a statistical method to *quantify* the magnitude of the distortions in quality signals that arise from non-representative attendance. Third, and most importantly, we estimate the *effects* of distortions on school choice and the allocation of public programs. This is the first paper to quantify the market consequences of distortions on educational systems. While we implement our analysis in the Chilean system, the implications of it go beyond both Chile and schooling. Multiple markets in which quality is imperfectly observed have quality disclosure systems, many of which may be prone to be distorted (Dranove and Jin, 2010). Moreover, whenever quality signals generated by the disclosure system feed into consumer and government choices, implications similar to those discussed in this paper might arise.[4]

This study relates to at least two literatures. First, the paper related to a literature that documents distortions in high-stakes testing. Distortions arise due to a number of reasons including diversion of resources, cheating, or manipulation of conditions under which tests are taken (see Figlio and Getzler 2002; Jacob and Levitt 2003; Jacob 2005; Figlio and Winicki 2005; Figlio 2006; Cullen and Reback 2006; Neal and Schanzenbach 2010; Apperson et al. 2016; Deming et al. 2016; Diamond and Persson 2017; Feigenberg et al. 2018; Quezada-Hofflinger and Von Hippel 2018; Dee et al. 2019, among others). In addition, non-behavioral factors such as mean reversion and random variation in the conditions under which tests are applied can also create distortions (see Kane and Staiger 2002; Chay et al. 2005; Ebenstein et al. 2016; Graff Zivin et al. 2018, among others). We provide evidence that non-representative test day attendance (regardless of how much of it originates in behavioral responses to incentives) is an additional source of distortions and compute its effects on school quality metrics.

Additionally, this paper contributes to the school choice literature. Several authors have shown that fees, distance between home and school, and school quality are the most relevant attributes for school choice (see Gallego and Hernando 2009; Feigenberg 2017, Neilson 2017 and Sánchez 2018 for Chile; Bayer et al. 2007; Hastings et al. 2009 and Walters 2017 for the U.S., among others). In addition, another set of studies investigates how information affects school choice, yielding mixed results (Hastings and Weinstein, 2008; Jensen, 2010; Mizala and Urquiola, 2013; Andrabi et al., 2017).[5]

[3] In concurrent papers, Quezada-Hofflinger and Von Hippel (2018) and Sánchez (2019) provide complementary evidence for this channel in Chile. Their main finding is that attendance of low-performing students decreases on test days. In contrast, our results also show that attendance of high-performing students increases on test-days. Their analyses differ from ours in that they focus on how test day attendance distorts the evaluation of a voucher program introduced in 2008, similar to Feigenberg et al. (2018).

[4] Examples of such settings are when quality information is provided to patients for health provider choice or when hygiene information is provided to consumers for restaurant choice (Dranove et al., 2003; Jin and Leslie, 2003).

[5] More broadly, our work is related to the literature in industrial organization studying disclosure and advertising (see Dranove and Jin 2010 and Bagwell 2007 respectively for reviews). As mentioned above, work that analyzes the effects

Our paper emphasizes the importance of *accurate* information in a context in which consumers are actively choosing.[6]

The remainder of the paper is structured as follows. Section 2 describes school markets and public programs in Chile. Section 3 describes the data and shows that low-performing students are underrepresented on test days. Section 4 estimates distortions in quality signals and discusses their determinants. Section 5 estimates a school choice model and studies the choice and welfare implications of distorted quality signals. Section 6 shows that two large public programs are misallocated because of non-random attendance on test days. The final section concludes.

2 Institutional context

2.1 School markets

Our analysis focuses on the Chilean primary school market. After a market-oriented reform was implemented in 1980, education has been provided by a mixture of public, private voucher and non-voucher schools. Students can apply and attend any school in the system, although funding varies across school types. Public schools are fully funded by the government. Private voucher schools are privately managed, although eligible for receiving public funding through vouchers. They are allowed to charge fees to parents in the form of copayments, although vouchers are phased out on the basis of those. Private non-voucher schools are not eligible for public funding.

Over the last three decades, the private sector has steadily increased its market share. In 2013, public schools had 38 percent of all students, while private voucher and non-voucher schools enrolled 54 and 8 percent of students respectively (Ministry of Education, 2013).

2.2 Public programs

Throughout the paper, we will refer to different public programs that are part of the Chilean educational system. For convenience, we briefly describe them in the remainder of this section.

Students in the Chilean educational system are eligible for vouchers. Public funding is provided

of quality disclosure in educational markets is somewhat limited and has yielded mixed results. Our paper relates to the case in which advertising is *informative*. Moreover, following the distinction proposed by Nelson (1970), the fact that schooling is an *experience good* implies that quality is hardly verifiable ex-ante, further implying that information acquired from advertising might be particularly important. This paper adds to this literature by focusing on educational markets, where there is limited work from an advertising perspective, and by measuring the implications of *deceptive advertising*.

[6]Our approach to measure the welfare implications of distorted quality signals distinguishes between choice and experience utility (Bernheim and Rangel, 2009). Recent work on the role of information frictions for insurance choice has adopted this insight (Handel and Kolstad, 2015; Handel et al., 2019). We adopt it to study information frictions in school choice.

on a per student basis and is linked to student attendance. However, the amount covered by vouchers depends on the characteristics of both students and schools. The baseline voucher program has been in place since the 1980's reforms. During the period we study, the amount of this voucher has varied across schools according to whether they offer full school shifts (*Jornada Escolar Completa*, JEC). Figure A.1 displays the evolution of the amount covered by vouchers during the years included in our dataset. As it can be noted, the amount paid to schools offering JEC is larger than what other schools receive.

In 2008, the Preferential Educational Voucher (*Subvención Escolar Preferencial*, SEP) was enacted as a complementary voucher targeted towards low-income households. Eligibility for this program is determined mostly by household income: households in the lowest third of the income distribution or that participate in the main social program offered by the government (*Chile Solidario*) are eligible for SEP vouchers. Some of our analysis distinguishes between low- and high-income households, mutually exclusive groups defined by SEP eligibility. All public schools are eligible for SEP vouchers, while private voucher school must subscribe in order to become eligible. Subscribing to the SEP program involves additional commitments by schools including limits to fees they might charge and designing resource management plans. SEP vouchers vary according to two school characteristics, namely the share of their students eligible for the SEP voucher and changes in the school's academic performance. Figure A.1 displays the evolution of SEP vouchers through time since their inception. Finally, both the autonomy in spending and the renewal in funding provided by this program is attached to school test scores. For further detail on the SEP program, see Correa et al. (2014).

The National System of Quality Measurement (*Sistema de Medición de la Calidad de la Educación*, SIMCE) has existed since 1988 and gives national standardized tests on different subjects. Tests are implemented every year at the national level for a subset of grades – see Figure A.2 for the timeline of test implementation. Test scores from SIMCE are comparable across schools and years. Tests are implemented by third party personnel. Moreover, average test scores are publicly disclosed and strongly disseminated at the aggregate school level, but are never made available to the public at the student level. Finally, test scores are never disclosed individually to teachers or students.

The National Performance Evaluation System (*Sistema Nacional de Evaluación de Desempeño*, SNED) is a school performance evaluation system that takes the form of a tournament and provides awards to improved schools. SNED operates as follows: (i) groups of *homogeneous* schools are constructed, within which the contest is implemented; (ii) every two years, an index is computed at the school level, which considers academic performance and improvement and socioeconomic integration among other outcomes; (iii) schools are ranked within their groups according to the value of such index; and (iv) schools covering 25-35 percent[7] of the total enrollment of each group get a monetary prize equivalent to around 80 percent of a teacher's monthly wage for each teacher in the school. Importantly,

[7] The coverage of the prize was increased to 35 percent of the enrollment of the group since 2006.

SIMCE test scores account for as much as 70 percent of the weight of the components used for the calculation of the SNED index (Contreras and Rau, 2012).

The Educational Traffic Lights program (*Semáforo Educacional*, ETL) was announced in April, 2010 and consisted of sending information to all households about local schools. That information included both test scores and a classification of schools as red, yellow or green according to their test scores, with clear cutoffs determining this outcome. An evaluation of this policy by Allende (2012) that uses the discontinuities in such classification for identification, finds that it effectively impacted school enrollment: households in the margin responded by enrolling more in yellow than red schools and more in green than yellow schools.

3 Data and attendance on test days

We use four administrative datasets provided by the Ministry of Education. First, is the record of schools operating between 2005 and 2013, in which we observe school type (public, private-voucher, private non-voucher), enrollment, fees, participation in government programs, and school addresses, which we use to construct markets. Second, we use student records between 2005 and 2013 (approximately 3.5 million per year), in which we observe enrollment (school, grade, classroom) and annual average GPA. Third, we use daily school attendance in 2013 to study heterogeneity in attendance on test days across the distribution of potential SIMCE performance. We argue that such heterogeneity is the source of distortions in quality signals. Finally, we use students' performance at SIMCE test as a measure of observed school quality. We focus on 4*th* graders because they are tested every year in the period 2005–2013 and because all schools offering 4*th* grade also offer 1*st* grade, the most relevant margin for school choice.

The focus on test scores as quality signals is appropriate given their contextual relevance. There is an extensive literature studying test scores and value added as quality measures for accountability systems (Meghir and Rivkin, 2011; Figlio and Loeb, 2011). In Chile, however, media outlets and government authorities use *test scores* as quality signals (McEwan et al., 2008) and survey evidence suggests that parents consider test scores important (Centro de Investigación y Desarrollo de la Educación, 2010). Accordingly, evidence shows that test scores affect school choice (Gallego and Hernando, 2009; Chumacero et al., 2011; Gómez et al., 2012; Neilson, 2017). In addition, the government uses these test scores to guide the allocation of public programs. Figure A.3 shows how test scores are publicly disseminated through media outlets, used for advertising by schools, and used as policy tools by the government.[8]

[8] The only measures of value added available for Chile are those computed by Neilson (2017). These value added measures are based on confidential administrative data. Figure A.4 displays the relationship between that measure of value added and test scores, which is positive and strong.

3.1 Descriptive statistics

Using the previously described administrative records, we construct two datasets: (i) a panel of schools, and (ii) a panel of students. Although the former includes all schools operating in the period 2005–2013, the latter is only available for public and voucher schools in 2013, which represent 93 percent of enrollment that year.

The school level dataset contains annual information on schools offering 4*th* grade in urban areas. The entry and exit of schools makes this panel unbalanced. There are 5,386 different schools and, on average, 4,640 schools operating in a given year. Table 1-A presents summary statistics for these schools: 39 percent are public, 52 percent are voucher schools, and 9 percent are private. The average school has approximately 50 students in 4*th* grade. More than half of schools charge no fees, and the average monthly fee is approximately \$48.[9] The average school test score is 255 and the standard deviation is 27.7.

Table 1-B presents descriptive statistics for the student level dataset. Students' academic performance is measured by their GPA, which ranges from 1 to 7, with a threshold of 4 as passing grade. The mean of this variable is 5.9. The last two variables are attendance rates on test and non-test days. The former is simply the average of two indicator variables that take the value of one if a student went to school on test days; there are two test days, so this variable has the value of 0, 0.5, or 1 at the student level. The latter is the average attendance in the five non-test days previous to test ones.

3.2 Attendance on test days

Schools average test scores (i.e., quality signals) are distorted if attendance on test days is non-random. Although every year there is anecdotal evidence (in the press) of some schools discouraging low-performing students to attend school during test-days, there has to date been no rigorous assessment of whether this practice is widespread. In this section, we show how student attendance patterns change on test days. While the government encourages full attendance on test days, schools face incentives to encourage high-performing students to attend and discourage low-performing students to do so. Therefore, it is not a priori clear what to expect. Our interest is not focused, however, on the average change in attendance, but rather on the *heterogeneity* behind this average change, both within and across schools.

In order to estimate the average change in students' attendance on test days, we compare the daily attendance rate of 4*th* graders ($\overline{A}_{4t}$, who take the test) to the daily attendance of 3*rd* graders ($\overline{A}_{3t}$, who

[9] All monetary units in the paper have been properly deflated and are measured in U.S. dollars using the early 2012 exchange rate.

do not take the test) around test days in 2013 (October 8*th* and 9*th*):

$$\Delta\overline{A} \;=\; \left(\overline{A}_{4T} - \overline{A}_{3T}\right) \;-\; \left(\overline{A}_{4\tau} - \overline{A}_{3\tau}\right) \tag{1}$$

where $t = T$ represents the two test days, and $t = \tau$ represent other days around test days. We calculate $\Delta\overline{A}$ in four subsamples of students: high-performing, above the 90*th* and 75*th* percentile of the GPA distribution; and low-performing, below the 10*th* and 25*th* percentile of the GPA distribution. In addition, to study the heterogeneity behind $\Delta\overline{A}$, we calculate the following school-specific changes in attendance on test day:

$$\Delta\overline{A}_j \;=\; \left(\overline{A}_{j4T} - \overline{A}_{j3T}\right) \;-\; \left(\overline{A}_{j4\tau} - \overline{A}_{j3\tau}\right) \tag{2}$$

where $\overline{A}_{jkt}$ is the average attendance rate of k*th* graders in school j and day t. The next section shows how a larger variance in $\overline{A}_j$ translates into more distorted quality signals.

Figure 1 displays $\Delta\overline{A}$ and $\Delta\overline{A}_j$. Panel (a) plots the differential attendance rate around test days. On average, attendance increases by 2 percentage points on test days. However, behind this increase there is relevant heterogeneity regarding *who* are the students attending on test days. The same figure shows that low-performing students are 2 percentage points more likely to stay at home on test days, whereas high-performing students are almost 4 percentage points more likely to attend school those days. The former is consistent with recent work by Quezada-Hofflinger and Von Hippel (2018) and Sánchez (2019), but the latter mechanism has not been documented in previous work.[10] Taken together, these results reveal that low-performing students are underrepresented on test days. However, these averages across schools mask significant heterogeneity, as shown in panel (b) by the distribution of $\Delta\overline{A}_j$ (the vertical line denotes the average increase of 2 percentage points).[11]

Overall, these patterns suggest that heterogeneity in attendance on test days is not the result of statistical noise, but rather that some behavioral response to the accountability system is in place. There are several potential explanations behind these patterns. On the one hand, this pattern might be driven by school behavioral responses to the incentives they face. On the other hand, this pattern may reflect that high- and low-performing students change their attendance decisions on test days for reasons unrelated to behavioral responses, e.g. differential stress/motivation from taking the test. These explanations need not be mutually exclusive. We explore the drivers of these patterns below. In any case, regardless of the drivers of these patterns of attendance on test days, the pattern itself

[10] A difference in our approach is the use of *daily* attendance data around test days for test-takers and non-takers, instead of only test-day attendance for test-takers. This allows us to study detailed patterns of heterogeneity in changes on attendance patterns around test days throughout the distribution of academic performance.

[11] We repeated this exercise using April 23th as a fake test day. We chose this day to maximize the number of school days before and after while remaining away in time from the actual test days. Figure A.5 shows that there is no differential attendance pattern across 4*th* and 3*rd* grades and the distribution of $\Delta\overline{A}_j$ is symmetric and centered around zero. This evidence further suggests that attendance on test day is a source of non-random distortions in quality signals.

creates distortions in school test scores. Importantly, the fact that this pattern is heterogeneous *across* schools causes observed quality signals in the educational market to be distorted in ways that affect school choice and the allocation of public programs. We now move on to estimate such distortions before quantifying the marketwide consequences.

4 Distortions in quality signals

4.1 Estimating undistorted quality signals

Quality signals are *undistorted* if all or a random sample of students take the test.[12] However, the patterns described in section 3.2 suggest that absenteeism on test days is not random. The empirical challenge to recover undistorted quality signals consists in estimating test scores for absent students.[13] If we can recover missing test scores, we can estimate undistorted quality signals that would be equivalent to the signals in a world with full or random attendance on test day. Our strategy to estimate missing test scores consists in using the multiple imputation methods developed by Rubin (1987). Using this strategy, we construct a panel dataset of distortions in quality signals for 2005–2013.

Estimating missing test scores. We begin by estimating the missing test scores. Let q_{ijt} be the test score of student i in school j and year t, and x_{ijt} be a vector of variables that predict test scores at the student level and that we observe for *all* students. Then, we estimate the following equation in the sample of test takers for each school in our dataset:

$$q_{ijt} \quad = \quad f(x_{ijt};\gamma_j) + \lambda_{jt} + \eta_{ijt} \tag{3}$$

where γ_j is a school specific vector of parameters, λ_{jt} are school-year fixed effects, and η_{ijt} is a mean zero random error term. Importantly, the vector x_{ijt} needs to contain strong predictors of test scores and be available for *all* students. We choose GPA and the following indicator variables: students who were in 4*th* grade the previous year and students who studied at a different school the previous year. Unsurprisingly, GPA is the strongest predictor of test scores at the student level, as displayed by Figure A.6. Moreover, given the quadratic empirical relationship between test scores and GPA, we include this variable as a quadratic polynomial. Note that equation (3) allows for the gradient of test scores to covariates in x_{ijt} to vary across schools. There are 7,500 schools in our dataset with, on

[12]We acknowledge that school quality signals are not affected only by the pool of test takers. Schools could also, for instance, affect students' effort during tests. But effort is unobservable our context, so here we define true school quality as the one that would be observed with the full population of students taking the test.

[13]Although daily attendance is not available for all years, it is possible to identify absenteeism on test days at the student level using the administrative records of annual academic performance and test scores: students with academic performance data but without test scores were absent on test days.

average, 270 test takers between 2005 and 2013. This means that our imputation method relies on 7,500 regression equations that use on average 270 observations and that we estimate using OLS.

We use equation (3) to predict test scores for absent students in the period 2005–2013. In order to account for the uncertainty related to the estimation of missing test scores, we estimate these test scores multiple times by drawing from the asymptotic variance of the estimated parameters $\hat{\gamma}_j$, an approach similar to that in Mas and Moretti (2009).[14] More precisely, for each absent student in our dataset, we generate one hundred estimated test scores based on equation (3), generating more than 20 million individual predicted test scores in the period 2005-2013.

Estimating distortions. After estimating test scores of absent students, we estimate "undistorted" quality signals using a simple simulation estimator. Let $\tilde{q}_{jt}^{(n)}$ be the average test score of school j in year t calculated using draw $n = 1, \ldots, 100$. Then, our estimate for an undistorted quality signal is:

$$\tilde{q}_{jt} = \frac{1}{100} \sum_{n=1}^{100} \tilde{q}_{jt}^{(n)}$$

The uncertainty of our estimates corresponds to the variance of the imputations $\tilde{q}_{jt}^{(n)}$. We order $\tilde{q}_{jt}^{(n)}$ from lowest to highest within a school and take the percentiles 2.5 and 97.5 to generate a 95 percent confidence interval for our estimate $\tilde{q}_{jt}$. We define distortions in quality signals as $\psi_{jt} \equiv q_{jt} - \tilde{q}_{jt}$, where q_{jt} is the observed (distorted) quality signal of school j in year t. Thus, a school with a positive distortion is one that signaled a higher quality than its true quality through its test score. Each distortion in our dataset has an associated distribution and a corresponding confidence interval.[15]

4.2 Descriptive statistics of distortions

The average distortion has a value of 2.7 test score points. Table A.1 presents descriptive statistics for different tested subjects. The distribution of distortions is remarkably similar across subjects as displayed by Figure A.9. Moreover, the correlation of distortions is high, as documented by Figure A.10. In what follows, we use the average of distortions across math and language in 4*th* grade, which were taken during all years in our sample.[16]

The average distortion is equivalent to 0.10 standard deviations (σ) of test scores at the school

[14] Alternatively, we could use a bootstrap procedure. We have done this as a robustness check and results are similar. We provide more details about draws from the asymptotic variance in Appendix A.1.

[15] Appendix A.2 shows that the estimated distortions are robust to a wide variety of econometric exercises. In particular, these are similar when using alternative specifications of equation (3) and similar when accounting for selective attendance on test day using a Heckman-selection model. In addition, this appendix also shows that the estimated distortions are empirically independent of the measurement error in individual test scores.

[16] Although the math and language tests are of higher-stakes than the natural and social sciences tests, these are taken by students in the same day, so we cannot exploit differences in stakes across subjects.

level. To assess their relative relevance, we compare the size of distortions to the impact of educational policies. Bellei (2009) evaluates a program that substantially extended school days in public and voucher schools in Chile and finds an impact of 0.06σ on test scores. Contreras and Rau (2012) find that the impact of SNED on test scores was between 0.14σ and 0.25σ. More broadly, Kremer and Holla (2009) and Glewwe and Muralidharan (2016) review multiple educational interventions in developing countries and a significant share display impacts smaller that 0.20σ. Similarly, a survey of field experiments in developed countries by Fryer (2016) finds that average treatment effects from school-based interventions are between 0.05σ and 0.07σ. Then, distortions in quality signals are of a relevant economic magnitude.

Figure 2-a presents estimated distortions for all schools in our data set. The y-axis represents distortions (in test score points), while the x-axis orders schools from lowest to highest distortion. In addition, distortions in green (gray) are (not) statistically different from zero. Approximately 31 percent of distortions are *statistically* larger than zero, and 80 percent of schools have a positive distortion. Figure 2-b presents the distribution of distortions. That (i) the average distortion is different from zero, and (ii) the distribution is not normal, make it clear that distortions in quality signals are not random variation in test scores. Moreover, we should note that *relative*, not *absolute*, distortions are relevant in terms of their potential implications. Figure A.13 presents an empirical analysis of the rank correlation between undistorted and distorted quality signals at the market-year level (see appendix C.2 for details on market definition). Approximately 60 percent of rank correlations are different from one, which suggests distortions in quality signals cause changes in the rankings of schools. Moreover, Figure A.14 shows that there were ranking changes in almost all large markets and in a sizable share of small markets.

Finally, we relate the estimated distortions with the motivating evidence presented in section 3.2. We would expect schools with higher differential changes in attendance in test days for high performing students (i.e. the difference between $\Delta\overline{A}_j^{high}$ and $\Delta\overline{A}_j^{low}$) to display larger distortions in quality signals. In this line, we start by calculating the difference in $\Delta\overline{A}_j$ between students above the 75*th* percentile and below the 25*th* percentile of the school's GPA distribution. Then, we study the relationship between this measure and our estimated distortions, displayed in Figure 3. Schools with the largest increases in relative attendance of high with respect to low ones on test days are also on average those with the highest estimated distortions, which provides evidence for our methodology for estimating distortions in quality signals.

4.3 Understanding distortions

What explains the variation in distortions in quality signals? As mentioned above, individual test scores are never disclosed to schools or students, and therefore we can rule out incentives for students (e.g. rewards, punishments) as drivers. We present a discussion of the determinants of distortions, in which we focus on a variety of school level characteristics and incentives as potential drivers of them. For this, we employ the panel dataset of distortions at the school level between 2005 and 2013.

4.3.1 Schools' characteristics

Fixed school characteristics. A significant share of the variation in distortions is explained by school time-invariant characteristics. If we regress distortions on school indicators, we can explain 36 percent of the variance. If we restrict attention to schools with statistically positive distortions, we can explain 60 percent of the variance. These percentages are large, especially considering that the maximum variation that can be explained is probably lower than one due to measurement error in the dependent variable. Which characteristics of schools predict distortions? Consider the following regression:

$$\psi_{jt} \quad = \quad X'_{jt}\theta + \nu_{mt} + \varepsilon_{jmt}$$

where X_{jt} is a vector of school attributes in year t and ν_{mt} is a market-year fixed effect. Markets are defined as isolated groups of schools, i.e., with no schools closer than 3 miles as discussed in appendix C.2. In order to account for the uncertainty in ψ_{jmt}, we present estimates weighted by (the inverse of) the 95 percent distortion confidence interval, thus accounting for the uncertainty associated to each distortion. Results are presented in Table 2-A and show that distortions are larger in small public schools, for-profit schools, schools serving relative low-income households, and schools with low attendance rates. These correlations are larger in schools with distortions that are statistically different from zero. Additionally, Table 2-B presents the autocorrelation of distortions, which is always positive and statistically different from zero. This positive autocorrelation serves as additional evidence that distortions are non-random but rather associated to school characteristics.

Time-varying school characteristics. We study whether variation in distortions can be explained by within-school-variation in observable characteristics including school fees, socioeconomic composition, undistorted quality, and measures of attendance and class size. In particular, we estimate:

$$\psi_{jt} \quad = \quad \beta X_{jt} + \eta_j + \nu_t + \varepsilon_{jt} \tag{4}$$

where X_{jt} is the covariate of interest, and η_j and ν_t are school and time fixed effects. Figure A.15 shows no relationship between any of these variables and distortions.[17]

[17]The only clear relationship is that between the number of students missing on test days and the magnitude of the distortion, which is positive as expected: missing students are a necessary condition for this distortion.

4.3.2 Competitive environment

The above results suggest that a substantive part of the distortions is explained by schools' fixed characteristics. Now we study whether part of the distortions can be explained by strategic behavioral responses. The first idea to explore is whether larger distortions are associated with the incentives that market environment creates for schools to signal higher quality (Shleifer, 2004). The market-oriented nature of the system suggests that schools facing more competition might choose to increase their quality signals using distortions. Dorfman and Steiner (1954) provide a useful framework to study firm behavior in contexts in which price and quality are jointly determined. The authors show that firms offer higher quality when facing more quality elastic consumers.[18] This section tests for this "quality elasticity" and related hypotheses.

We exploit within school variation in variables related to the competitive environment. We proceed by estimating regressions following equation (4). The variables we consider include the number of schools in the market, average quality, fees and distortion of rivals, and the position of a school in the distribution of fees and quality in the market. We also employ the estimates from our school choice model in section 5 to calculate quality demand elasticities.

Figure A.16 displays results graphically. Although changes in the number of schools in the market and changes in average attributes of competitors are uncorrelated with distortions, demand quality elasticity is strongly correlated with distortions. The latter result is consistent with Dorfman and Steiner (1954): schools facing higher quality elasticity optimally choose to signal higher quality. This result is reinforced by the fact that schools in higher percentiles of the market-level quality distribution also seem to introduce higher distortions.

4.3.3 School and teacher incentives

Potential gains in test scores. The extent to which schools may increase test scores through differential attendance on test days could be one of the drivers of distortions. We construct a measure of potential gains from non-random attendance on test days by comparing predicted school test scores a school would obtain if the ten percent of students in the bottom of the GPA distribution was absent on test day with the predicted school test scores if all students attended such day. Figure A.17 displays the correlation between estimated distortions and this measure of potential gains, which is strong

[18] Dorfman and Steiner (1954) analyze the behavior of a monopolist and argue that quality is optimally set following the condition:

$$q = \frac{p}{c_q}\frac{\eta^q}{\eta^p}$$

where q is quality, p is price, c_q is the cost of quality, and η^q and η^p are the quality and price demand elasticities, respectively. In our interpretation, however, we use their result to approximate the case of imperfect competition with multiple firms and the analysis of a particular firm facing residual demand which is one way of modeling school behavior in this market setting (Neilson, 2017). In our setting, we argue that observed quality q can be increased by either increasing true quality or introducing higher distortions.

and positive: schools that gain more from non-random attendance on test days also display higher distortions.

The role of public programs. Finally, we exploit quasi-experimental variation from three government programs to understand the role of perverse incentives created by public programs as drivers of distortions. This is in line with the literature that studies school behavioral responses to accountability systems (Figlio and Getzler, 2002; Kane and Staiger, 2002).

We start by analyzing incentives placed by the SEP voucher program. This program placed incentives for school to increase test scores through tying both autonomy in expenditure and the renewal of government funding to SIMCE test scores.[19] We exploit variation in the timing of adoption of the SEP program to implement an event study analysis and examine the relationship between incentives placed by the SEP program and distortions in test scores. Appendix B presents details of the analysis and Figure A.18 results. We find that schools that adopt the SEP program display a significant increase of 0.7 points (0.17σ) in distortions in years after adoption relative to previous years, which persists four years after adopting the program. This pattern suggests that pressures for obtaining government funding tied to test scores might be inducing schools to engage in gaming-like behaviors to raise test scores.[20] We interpret these results with caution, because adoption of the SEP program is a choice of schools (in particular, for voucher schools) and therefore program adoption could be correlated with school unobservables also driving distortions. However, the well behaved pre-trend in effects leading to the event of adoption in Figure A.18 limits this concern.

The second government program we study is the SNED teacher incentives program which effectively increases teachers' wages if students in the school obtain high test scores, providing variation in incentives depending on the probability that a school will earn the prize (Contreras and Rau, 2012). Schools, and teachers in particular, might react to the likelihood of obtaining these prizes by affecting student attendance patterns on test days in order to increase the school's average test score.[21] If anything, we would hypothesize that schools closer to the prize threshold would display larger distortions. However, our results show that distortions are not higher in schools that are more likely to

[19]In particular, incentives placed by the SEP program operate through a classification of schools that is based largely on SIMCE test scores. This classification then determines (i) the degree of autonomy that schools are provided in spending government funding offered by the program, such that schools with higher test scores have more flexibility than those with lower test scores; and (ii) the renewal of the affiliation of schools to the program after four years in it also depends on SIMCE test scores. See Correa et al. (2014) for further details.

[20]This evidence is consistent with findings in recent work by Feigenberg et al. (2018), Quezada-Hofflinger and Von Hippel (2018), and Sánchez (2019). All of these papers find that part of the increase in test scores induced by the program can be explained by changes in test day attendance.

[21]A condition for this test to be able to capture school behavioral responses is that schools are informed about the likelihood they obtain the monetary incentives provided by the SNED program. The evidence in Contreras and Rau (2012) suggests this condition holds in our setting. However, in informal discussions, personnel of the Agency for the Quality of Education suggested that schools might not be well informed about such likelihood. Thus, the results of this test should be interpreted with caution.

win the prize, providing some evidence against the hypothesis that teachers manipulate attendance to increase test scores. See Appendix B and Figure A.19 for more details and results.

The third government program we exploit is the ETL information program. This program classified schools according to test scores: red (bad), yellow (regular), and green (good). This information was disseminated to households in order to aid school choice. The cutoffs of these categories provide quasi-experimental variation in the incentives to manipulate test scores.[22] We find some evidence that low quality schools have higher distortions around the cutoff between red and yellow schools, but no differential distortions in the cutoff between yellow and green schools. Moreover, these differences mostly disappear once school characteristics are controlled for. Overall, these patterns do not provide evidence for this mechanism being the main driver of distortions. See Appendix B and Figure A.20 for further details and results.

4.3.4 Discussion

The empirical patterns presented in this section improve our understanding of distortions in several dimensions. First, distortions are a non-random phenomenon. Second, fixed school characteristics explain a large share of the variation in distortions.[23] Third, we rule out that distortions are driven by within-school variation in a number of relevant observable school characteristics. Fourth, we provide suggestive evidence that the market environment is correlated with distortions through the quality demand elasticity that schools face. Fifth, we provide mixed evidence for the effects of incentives induced by public programs driving variation in distortions: government funding attached to test scores displays a strong positive relationship with distortions, while both teacher monetary incentives and quality disclosure policies display no relationship with them. The last two sets of results suggest that at least part of the distortions in quality signals are driven by strategic school behavioral responses to market and government incentives.

As mentioned above, regardless of the factors driving distortions, we can estimate the consequences of the distortions both on school choice and on the allocation of public programs.

---

[22]The timing for this exercise is relevant. The SIMCE test was taken shortly after ETL report cards had been distributed to households, in the same academic year. Therefore, any distortions schools could have introduced in October, 2010 would affect test scores before any households reactions in terms of school choice.

[23]A plausible hypothesis behind this result is the existence of different school *unobserved types*, where some types choose to distort test scores and some choose not to do so. These types might be related to, for example, school principals, who we do not observe. This hypothesis is consistent with either the strategic behavioral response hypothesis (e.g. the types of schools that *do* choose to distort would do so more strongly when facing more incentives to do so) or the non-strategic one (i.e. the observed distortion might be in part explained by schools' efforts to manipulate attendance and in part by the students' own reaction to a test).

5 Implications for school choice

What is the impact of distortions in quality signals on school choices by households? The school choice literature has shown that households consider school quality when choosing schools (e.g. Gallego and Hernando 2009 and Neilson 2017 for Chile). In our setting, school quality signals are distorted and therefore households are choosing schools on the basis of misperceived attributes. The value that households ultimately obtain from a school is the true quality of their school choice.[24] Distortions in quality signals might therefore affect household school choices and welfare. In this section, we quantify such effects by using an estimated school choice model to study a counterfactual exercise of providing households with *undistorted* quality signals.

We develop and estimate a discrete choice school choice model, which is tailored to the Chilean institutional framework and based on recent work on school choice (Bayer et al., 2007; Neilson, 2017). In order to focus on the counterfactual analysis, we leave the discussion about the model, its identification and estimation to Appendix C. In the model, households are indexed by i and schools are indexed by j. Households choose the school that maximizes their utility among schools available in the market. In particular, the indirect utility of household i of type r from enrolling their child in school j is:

$$u_{ij}^r = \sum_k x_{k,j}\beta_k^r + \xi_j^r + \beta_d^r d_{ij} + \varepsilon_{ij} \tag{5}$$

where x_j^r are the attributes of school j, including fees p_j^r and quality q_j, among others; ξ_j^r are school characteristics that are unobserved to the econometrician; d_{ij} is distance to school j; and ε_{ij} is an idiosyncratic preference shock. The model allows for observed heterogeneity in preferences across low- and high-income households.

Our estimates of preferences over schools are in line with previous work on school choice (e.g. Gallego and Hernando 2009, Hastings et al. 2009, and Neilson 2017). We exploit detailed data on school and household attributes and geocoded locations for almost 100,000 students and more than 1,500 schools.[25] In summary, our estimates imply that households prefer schools with lower fees, higher quality, and that are close to their residence. Furthermore, there are interesting patterns of heterogeneity across low- and high-income households: low-income households are more price-sensitive, less quality-sensitive and more distance-sensitive. This heterogeneity in preferences implies that high-income households' willingness to pay for quality is three times higher than that of low-income

[24]This is related to the distinction between *choice* utility and *experience* utility stated by Bernheim and Rangel (2009), by which some elements of the choice environment may be relevant for constructing positive descriptions of choice behavior, but not for welfare analysis. Throughout this section, we emphasize this aspect and account for it when measuring implications of distorted quality signals.

[25]We discuss estimation, market definition and results in detail in Appendices C.1, C.2 and C.3, respectively.

households. This pattern will drive substantial heterogeneity across households in the counterfactual analysis we develop below.

In order to compute the effects of distorted quality signals on choices and welfare, we define two scenarios: *baseline* and *counterfactual*. The former is the environment in which households actually choose schools. The latter is a counterfactual environment in which households are provided with undistorted information about school quality. This exercise rules-out changes in other variables (e.g. school fees and school investments) as well as the existence of capacity constraints. While those might be relevant margins of supply side behavior in this market, we argue that the impacts of the policy we evaluate in this counterfactual exercise would induce small equilibrium responses by schools.

Throughout this section, we utilize our estimated school choice model and the observed vector of school characteristics X_j to compute choice probabilities and consumer welfare for the baseline scenario. For the counterfactual scenario, calculations use a counterfactual vector of school characteristics $\tilde{X}_j$, which includes undistorted school quality in place of observed quality signals.[26]

5.1 Effects on school choice

We begin the analysis by examining school choice probabilities across both scenarios, with details presented in appendix C.4. Figure A.24 displays changes in choice probabilities. Despite the fact that the magnitude of estimated distortions is moderate, there is significant heterogeneity.[27] This shows that changes in the quality disclosure system would induce changes in households' choices. However, given that households have a limited number of schools in their choice sets, these changes in choice probabilities might only induce actual changes for a small fraction of households. Note that high-income households display more variation in the computed changes, which is driven by their higher quality sensitivity. This stands in contrast with potential gains from the policy, as the average distortion in low-income household choice sets are 0.33σ higher than those in high-income household choice sets. Despite that difference, a simple simulation based on the proposed model and our estimates shows that 3.3 percent of low-income households and 3 percent of high-income households would be induced to change their school choice when provided undistorted quality information. The higher willingness to pay for quality of high-income households explains these similar switching rates despite the large gap in distortions faced by both groups. We denote this subpopulation as *switchers*.

The attributes of schools that households choose differ across scenarios, as displayed by Table 3. We report both the average across switchers and across all households. Columns 1 and 3 in Table 3 display

[26] All calculations using the estimated model utilize the results for the second stage from our preferred specifications: columns 6 and 9 of Table A.4.

[27] This pattern holds when restricting the analysis to the set of schools actually chosen by parents as displayed by Figures A.24-C and A.24-D.

results for switchers. First, note that in the baseline scenario, switchers were receiving substantially less quality than the average household, which suggests that switchers mainly had chosen schools with highly distorted quality signals. Conditional on switching, we observe that households are willing to travel longer distances to chosen schools, to pay higher fees and, importantly, that they choose schools with remarkably higher true quality. In particular, our results show that low- (high-) income switchers would choose schools with 0.71σ (0.74σ) higher true quality in the counterfactual than the baseline scenario. This would be coupled by an increase in fees paid to chosen schools of 0.20σ (0.49σ) for low-income (high-income) switchers and, similarly, an increase in distance travelled to chosen schools of 0.04σ (0.05σ). Switchers thus change their choices in a substantial way when provided with undistorted information on school quality.

Columns 2 and 4 in Table 3 display results for the average across all households. For this population, changes in predicted distance to chosen schools and fees are small. This result is as expected, since non-switching households are unaffected by the policy we evaluate. The average changes in attributes of chosen schools by low- and high-income households are not larger than 0.03σ for any of the attributes considered.

5.2 Effects on household welfare

We now calculate the effects of providing undistorted quality signals on consumer surplus. In the baseline scenario, households choose schools using observed school quality, which, as discussed, is distorted. However, consumers' effective utility is determined by undistorted school quality. Thus, our baseline scenario is a case in which *choice* utility and *experience* utility differ (Bernheim and Rangel, 2009). This is not the case in the counterfactual scenario, in which both coincide.

Let u_{ij} be the utility of household i from school j under distorted school quality, choice utility. Similarly, let $\tilde{u}_{ij}$ be the utility of household i from school j under undistorted school quality, experience utility. These two utilities are related in our setting. Given that the only difference between choice and experience utility is the misperception of quality under the former, we know that $\tilde{u}_{ij} = u_{ij} + \tau_j$, where τ_j measures the wedge between choice and experienced utility from school j. Under the utility function assumed in equation (5), we know that $\tau_j = \beta_q(\tilde{q}_j - q_j)$.[28]

The choices household i would make in each scenario would be:

$$\begin{aligned} j_i^* &= \arg\max_j \{u_{ij}\}_{j \in \mathcal{J}_i} \\ \tilde{j}_i^* &= \arg\max_j \{\tilde{u}_{ij}\}_{j \in \mathcal{J}_i} \end{aligned}$$

[28]These linear relationships between observed and true quality and between choice and experience utility are similar to those analyzed in Train (2015). From this expression for τ_j, it becomes clear that at baseline all schools with positive distortions have $\tau_j < 0$, such that experience utility from those schools is lower than choice utility from them.

which might or might not differ. Importantly, if the choice is the same in both scenarios then there is no welfare loss from distorted quality signals for household i, as experience utility is the same in both cases. This makes it clear that welfare losses will be driven by households that were changing their behavior due to distorted quality signals.

The change in household welfare from providing undistorted information would therefore be the difference in experience utility between the counterfactual and baseline scenarios, $\tilde{u}_{i\tilde{j}^*} - \tilde{u}_{ij^*}$. Using the fact that $\tilde{u}_{ij^*} = u_{ij^*} + \tau_{j^*}$, we can compute the expected monthly change in consumer surplus as:

$$E[\Delta CS_i] = \frac{1}{\beta_p}\left[\log \sum_j \exp(\tilde{v}_{ij}) - \log \sum_j \exp(v_{ij}) - \sum_j P_{ij}\tau_j\right] \tag{6}$$

where we define $\tilde{v}_{ij} \equiv \tilde{\delta}_j + \beta_d d_{ij}$ and $v_{ij} \equiv \delta_j + \beta_d d_{ij}$ for notational simplicity. The first and second terms measure consumer surplus under undistorted and distorted school quality information respectively, and the results follow from the inclusive value formula in Small and Rosen (1981) given the assumed utility function. The third term measures the expected difference between choice and experience utility at baseline, according to school choice probabilities P_{ij}. Dividing by β_p simply transforms the welfare loss to monetary units. Equation (6) calculates the average gain in consumer surplus across all households in the sample. We can then compute average gains in consumer surplus for switchers or aggregate these gains across different dimensions. These welfare gains can alternatively be interpreted as the average willingness to pay of households for undistorted quality information.

Results from welfare calculations are displayed by Table 4 and show that expected welfare would increase in the counterfactual scenario for all households. This is as expected: non-switchers will obtain the same welfare in both scenarios, whereas switchers will be strictly better off. The average yearly welfare gain for switchers is \$53 among low-income households and \$174 among high-income households. Gains for switchers are thus sizable: low-income (high-income) switchers would experiment welfare gains of 11 (36) percent of the average school fee in our sample. Average welfare gains across households are smaller. For low-income households, the average yearly welfare gain we estimate is \$1.7. The average yearly welfare gain for high-income households is \$5.3. Scaling up these results for the educational system, welfare gains would add up to \$7 million annually.[29]

5.3 Heterogeneity in welfare gains

The fact that high-income households benefit more than low-income households from the information policy is evident, and the magnitude of the differences is large. There are two potential explanations for this. First, the former are more quality-sensitive and less price and distance-sensitive than the

[29] Aggregate welfare gains are calculated as the average yearly welfare gain from undistorted information, multiplied by the total number of students between 1st and 8th grades in 2014, which was 1,939,926.

latter. Therefore, they will be more willing to take advantage of relative changes in perceived quality of schools in the market. Second, the spatial distribution of households and schools in the market differs systematically across low- and high-income households, giving them potentially differential opportunities to improve their choices in the counterfactual.

We can use our model and estimates to explore how heterogeneity in preferences and market opportunities determine the observed gap in welfare gains from disclosure of true quality. Results from these additional counterfactual calculations are displayed in Table 4. We start by studying how differences in preferences determine lower welfare gains for low-income households. First, we let low-income households be as quality-sensitive as high-income ones. The share of switchers among low-income households would increase by 0.8 percentage point to 4.1 percent, and the average yearly welfare gains for switchers would increase to \$101.[30] Second, we let low-income households have the same preferences as high-income households on all school attributes. The share of switchers increases by 0.6 percentage point to 3.8 percent. Average yearly gains for low-income switchers in this counterfactual would climb to \$181, more than three times those in the first counterfactual and higher than those for high-income switchers.[31] These results imply that differences in preferences are enough to explain the gap across groups in welfare gains from the proposed information policy. Moreover, they highlight the key role that households' quality-elasticity plays in determining the impacts of information policies for school choice.

Finally, we explore the role that the spatial distribution of schools and households play in explaining the gap in welfare gains across groups. We measure welfare gains from the evaluated policy for low-income households if they were located in the same place as high-income households. Our results show that average welfare gains in that setting would be essentially the same that we found in our baseline results above. The share of switchers in this case would be lower than in the first counterfactual, at 2.4 percent, while yearly welfare gains for low-income switchers would be only slightly larger than in such counterfactual, \$65. This result implies that, in our setting, differences in market opportunities faced by low- and high-income households play a minor role in explaining the gap in welfare gains from undistorted quality information.

[30] Recall that in conditional logit models, coefficients are normalized by the scale parameter of the idiosyncratic preference shock, σ_ε^r, which may vary across household types. Thus, in practice, this counterfactual is not exactly letting the low-income have the quality preference of the high-income, but rather the estimated normalized preference coefficient of such group. This is equivalent to making low-income households almost twice as price sensitive as estimated.

[31] The fact that welfare gains for the low-income when endowed with preferences of high-income households are larger than those when endowed with such preference only over school quality comes partly from the fact that we estimate high-income households to be less price-sensitive. This implies that the willingness to pay for a given increase in quality is higher than under low-income preferences as can be noted in equation (6).

5.4 Discussion

We have estimated a school choice model and studied a counterfactual exercise by which information on undistorted quality signals is provided to households. Results point in three directions. First, distortions in quality signals have effects on choices, as choice probabilities change in the counterfactual scenario. Second, households react to the change in the quality disclosure system mostly by increasing the probability of choosing higher quality schools. There is a shift of students towards relatively high quality schools available in the market. Third, our welfare calculations point towards sizable gains for switchers. Gains are larger for high-income households, which is driven by them being more quality-sensitive and less price-sensitive. Complementary policies that increase low-income households quality-sensitivity would increase welfare gains from this policy.

Throughout this section, we assumed that households are uninformed about distortions in quality signals. If they were informed, they would optimally incorporate that information and adjust their choices accordingly. Because calculating distortions is a complex task and all the necessary inputs to estimate them are unobserved by parents (e.g. test day attendance), we argue that parents are unlikely to incorporate them in their decisions. If households had partial knowledge about distortions, then welfare gains for switchers would certainly be lower and our estimates would be an upper bound.

The magnitude of welfare gains for switchers suggests that it might be socially beneficial to invest in quality disclosure systems that reduce distortions in educational markets. Note that our exercise does not evaluate the welfare effects of the disclosure system in place, but rather those of distorted quality signals given the current disclosure system. Moreover, these welfare calculations do not consider the social costs of potential hidden actions that might be driving distortions. In that sense, our results provide a lower bound for welfare gains from correcting distortions in this market.

6 Misallocation of public programs

There is a second set of implications of distorted quality signals. Multiple public programs are allocated using rules that follow directly from test scores. Thus, distortions in test scores will induce misallocation of funds and resources associated with these programs. This section quantifies such misallocation for two public programs: teacher bonuses and school choice information.

6.1 Teacher bonuses

As explained in section 2.2, teachers are awarded bonuses by the SNED program depending on their school's average test score. In 2012, the total amount of public resources transferred to schools in the form of teacher bonuses reached 15 million U.S. dollars. The sharp discontinuity to assigning resources is based on the following index for each school:

$$I_{jgt}(q_{j\tau}, q_{j\tau-1}, X_{j\tau}) \quad = \quad \omega_1 q_{j\tau} + \omega_2(q_{j\tau} - q_{j\tau-1}) + \boldsymbol{\omega}_3' \mathbf{X}_{j\tau} \tag{7}$$

where I_{jgt} is the index of school j, in group g, and year t; $q_{j\tau}$ is the average test score in year τ; $\mathbf{X}_{j\tau}$ is a vector of attributes; and $(\omega_1, \omega_2, \boldsymbol{\omega}_3)$ are weights chosen by the government, with $\omega_k \in (0, 1)$ and $\sum_k \omega_k = 1$. Note that: (i) $t > \tau$, otherwise the index cannot be computed as the inputs to calculate it are not observed, (ii) all input variables are mapped to the $[0, 1]$ interval before computing the index, and (iii) groups g are defined by the government using school attributes.

We say there is misallocation of public funds if teacher bonuses were given to schools that would not have receive bonuses in a counterfactual scenario without any distortions in quality signals. In particular, using our estimates for undistorted quality signals $(\tilde{q}_{j\tau}, \tilde{q}_{j\tau-1})$, we calculate schools undistorted indices using equation (7), $\tilde{I}_{jgt} = I_{jgt}(\tilde{q}_{j\tau}, \tilde{q}_{j\tau-1}, \mathbf{X}_{j\tau})$, and reallocate bonuses based on these undistorted measures.

Figures 4-a and 4-b present the actual and the counterfactual assignment of bonuses. To the left of the threshold (vertical line) are the schools that did not get bonuses, and to the right are the schools that did. The percentage of public resources that were misallocated is the total amount of money that was incorrectly given to some schools over the total amount of resources that schools received. We estimate that 13 percent of teacher bonuses were misallocated.

Although intuitive, our method to calculate misallocation of public resources still needs to account for the uncertainty associated with the estimation of undistorted quality signals. For this, recall that each school-year distortion has an associated distribution. We proceed by taking 1,000 independent draws of distortions from their school-year distribution – a normal distribution with a school-year specific mean and standard deviation – and calculate the percentage of misallocated public resources 1,000 times. Bounds for our misallocation estimates can be constructed using the estimated distribution of misallocation.

Our estimates indicate that 13 percent of teacher bonuses were delivered to the incorrect schools, which is equivalent to \$2 million every two years or approximately \$20 million since this public program started in 1996. This estimate is significantly different from zero and precise: we can rule out misallocation of public resources being less than 11 percent.

6.2 Information for school choice

A quality disclosure program was implemented in 2010 ("Educational Traffic Lights"), aimed at providing simpler information about school quality (more details in section 2). Schools were classified, based on the average test scores of 4*th* and 8*th* graders, into three mutually exclusive categories. Maps with school categories were disseminated across counties with the explicit objective of affecting parents information set.

Let $c_j = \{r, y, g\}$ be the category of school j (red, yellow, green). Schools were assigned to categories using the following formula:

$$c_j(q_{jt}) \quad = \quad r \cdot 1[q_{jt} < \underline{s}] + y \cdot 1[\underline{s} < q_{jt} < \bar{s}] + g \cdot 1[q_{jt} > \bar{s}] \tag{8}$$

where q_{jt} is the average test score of school j in year $t = 2009$, and $(\underline{s}, \bar{s})$ were thresholds decided by the government. These thresholds corresponded to one standard deviation lower ($\underline{s}$) and higher ($\bar{s}$) than the average test score of all schools.

Equation (8) makes it clear that the provided information is directly linked to distorted quality signals. Because the formula used to categorize schools is known, we can replace distorted quality signals by undistorted ones, assign undistorted categories $\tilde{c}_j = c_j(\tilde{q}_{jt})$, and calculate the percentage of schools that were incorrectly categorized. In order to account for the uncertainty in our undistorted quality signals, we follow the same strategy as in the previous section.

Figures 4-c and 4-d present our results. Our estimates indicate that approximately 4 percent of schools were assigned to an incorrect category. Moreover, we can rule out that fewer than 3 percent of schools were misassigned. Using the causal effects reported in Allende (2012) we calculate that, as a consequence of this misallocation of categories, approximately 5,000 students (two percent of the 1*st* grade cohort) attended schools in misallocated categories. The welfare implications for the compliers are, however, not straightforward to calculate as some children attended higher-quality and some attended lower-quality schools.

7 Conclusion

We have shown that significant distortions in quality signals are in place in the Chilean educational market, which is dependent on high-stakes testing. In particular, we have quantified how non-random attendance on test day causes school quality signals to be distorted. We study the determinants of these distortions and we find that they are largely explained by fixed school characteristics and that strategic behavioral responses by schools play a role. The latter is consistent with the so-called Campbell's Law: the higher the stakes are for an indicator of a social phenomenon, the more liable it is to be distorted (Campbell, 1979). Distortions, however, are not *per se* a reason of concern. To claim distortions have costs, we need to study the impacts they have on decisions. The Chilean market-oriented educational system is particularly interesting to study such impacts because test scores are not just used for the two objectives of quality assessment and performance evaluation emphasized by Neal (2013), but rather for three, as they also feed school choice. As we have shown that distortions have negative impacts on school choice and induce misallocation of public programs, we conclude that distortions can impose significant costs in educational markets.

Our study is, to the best of our knowledge, the first to quantify the *market* consequences from distortions in quality signals. Further research is required to quantify other distortions and to address other margins of educational markets. We highlight that the institutional environment might determine the magnitude and impacts of distortions.[32] Market-oriented educational systems such as the one we have studied –where test scores play a key role as quality signals in disclosure policies– might be particularly prone to exacerbating the consequences of distortions.

Our results have several policy implications. Previous work has emphasized the importance of providing information to parents, while our work emphasizes the importance of providing *undistorted* information. A simple solution within the current system is to calculate undistorted quality signals using the imputation method we have proposed or to report median test scores instead of averages. Both seem to be better solutions than requiring a minimum attendance rate (e.g., 95 percent in No Child Left Behind) in contexts where test scores can affect school choice. In addition, we emphasize that the magnitude of elasticities determines the extent to which households can benefit from information policies. In school markets, we argue that complementary policies that increase quality-sensitivity of low-income households might enable them to benefit more from accurate information. Finally, our results on misallocation of public programs provide an argument against sharp assignment rules for public programs based on variables prone to distortions. Multiple programs in different countries and sectors are assigned through such rules and might be subject to misallocation.

References

Alfaro, R. and Fuenzalida, M. (2009). Imputación múltiple en encuestas microeconómicas. *Cuadernos de Economía*, 46:273–288.

Allende, C. (2012). The impact of information on academic achievement and school choice: evidence from Chilean "Traffic lights". Manuscript.

Andrabi, T., Das, V., and Khwaja, A. (2017). Report cards: the impact of providing school and child test scores on education markets. *American Economic Review*.

Apperson, J., Bueno, C., and Sass, T. R. (2016). Do the cheated ever prosper? the long-run effects of test-score manipulation by teachers on student outcomes.

Bagwell, K. (2007). The economic analysis of advertising. In Armstrong, M. and Porter, R., editors, *Handbook of Industrial Organization*, volume 3, pages 1701–1844. Elsevier.

Bayer, P., Ferreira, F., and McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, 115(4):588–638.

[32] A relevant institutional dimension is the level of corruption. Interestingly, Chilean counties with higher levels of corruption have larger distortions in quality signals (see Tables A.8 and A.9). This suggests that our findings might be exacerbated in settings with different levels of corruption.

Bellei, C. (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.

Bernheim, B. D. and Rangel, A. (2009). Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *The Quarterly Journal of Economics*, 124(1):51.

Campbell, D. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2:67 – 90.

Cartwright, E. and Menezes, M. L. (2014). Cheating to win: dishonesty and the intensity of competition. *Economics Letters*, (122):55–58.

Centro de Investigación y Desarrollo de la Educación (2010). VIII encuesta a actores del sistema educativo 2010.

Chay, K. Y., McEwan, P. J., and Urquiola, M. (2005). The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review*, 95(4):1237–1258.

Chumacero, R. A., Gómez, D., and Paredes, R. (2011). I would walk 500 miles (if it paid): Vouchers and school choice in Chile. *Economics of Education Review*, 30:1103–1114.

Contreras, D. and Rau, T. (2012). Tournament incentives for teachers: evidence from a scaled-up intervention in Chile. *Economic Development and Cultural Change*, 61(1):219–246.

Correa, J. A., Parro, F., and Reyes, L. (2014). The effects of vouchers on school results: Evidence from Chile's targeted voucher program. *Journal of Human Capital*, 8(4):351–398.

Cullen, J. B. and Reback, R. (2006). Tinkering towards accolades: School gaming under a performance accountability system. In Gronberg, T. J. and Jansen, D. W., editors, *Improving School Accountability*, volume 14. Emerald Group Publishing Limited.

Dee, T. S., Dobbie, W., Jacob, B. A., and Rockoff, J. (2019). The causes and consequences of test score manipulation: Evidence from the New York regents examinations. *American Economic Journal: Applied Economics*.

Deming, D. J., Cohodes, S., Jennings, J., and Jencks, C. (2016). School accountability, postsecondary attainment and earnings. *Review of Economics and Statistics*.

Diamond, R. and Persson, P. (2017). The long-term consequences of teacher discretion in grading of high-stakes tests. Manuscript.

Dorfman, R. and Steiner, P. O. (1954). Optimal advertising and optimal quality. *The American Economic Review*, 44(5):826–836.

Dranove, D. and Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, 48(4):935–963.

Dranove, D., Kessler, D., McClellan, M., and Satterthwaite, M. (2003). Is more information better? the effects of report cards on health care providers. *Journal of Political Economy*, 111(3):555–588.

Ebenstein, A., Lavy, V., and Roth, S. (2016). The long run consequences of high-stakes examinations: Evidence from transitory variation in pollution. *American Economic Journal: Applied Economics*.

Feigenberg, B. (2017). Priced out: Aggregate income shocks and school pricing in the Chilean voucher market. Manuscript.

Feigenberg, B., Rivkin, S., and Yan, R. (2018). Illusory gains from Chile's targeted school voucher experiment. Working Paper 23178, National Bureau of Economic Research.

Figlio, D. and Loeb, S. (2011). School accountability. In Eric A. Hanushek, S. M. and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 3, pages 383–421. Elsevier.

Figlio, D. and Winicki, J. (2005). Food for thought: the effects of school accountability plans on school nutrition. *Journal of Public Economics*, 89(2-3):381–394.

Figlio, D. N. (2006). Testing, crime and punishment. *Journal of Public Economics*, 90(4-5):837–851.

Figlio, D. N. and Getzler, L. S. (2002). Accountability , ability and disability: Gaming the system. Working Paper 9307, National Bureau of Economic Research.

Fryer, R. (2016). The production of human capital in developed countries: Evidence from 196 randomized field experiments. volume 1 of *Handbook of Field Experiments*. Elsevier.

Gallego, F. and Hernando, A. (2009). School choice in chile: Looking at the demand side. *Pontificia Universidad Catolica de Chile Documento de Trabajo*, (356).

Gilpatric, S. M. (2011). Cheating in contests. *Economic Inquiry*, 49(4):1042–1053.

Glewwe, P. and Muralidharan, K. (2016). Chapter 10 - improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. volume 5 of *Handbook of the Economics of Education*, pages 653 – 743. Elsevier.

Gómez, D., Chumacero, R., and Paredes, R. (2012). School choice and information. *Estudios de Economía*, 39(2):143–157.

Graff Zivin, J., Hsiang, S., and Neidell, M. (2018). Temperature and Human Capital in the Short and Long Run. *Journal of the Association of Environmental and Resource Economists*, 5(1):77–105.

Handel, B. and Kolstad, J. (2015). Health insurance for"humans": Information frictions, plan choice, and consumer welfare. *American Economic Review*, 105(8):2449–2500.

Handel, B., Kolstad, J., and Spinnewijn, J. (2019). Information frictions and adverse selection: Policy interventions in health insurance markets. *The Review of Economics and Statistics*.

Hastings, J., Kane, T. J., and Staiger, D. O. (2009). Heterogeneous preferences and the efficacy of public school choice. Manuscript.

Hastings, J. S. and Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *Quarterly Journal of Economics*, 123(4):1373–1414.

Jacob, B. A. (2005). Accountability, incentives and behavior: the impact of high-stakes testing in the chicago public schools. *Journal of Public Economics*, 89:761 – 796.

Jacob, B. A. and Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics*, 118(3):843–877.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *Quarterly Journal of Economics*, 125(2):515–548.

Jin, G. Z. and Leslie, P. (2003). The effect of information on product quality: Evidence from restaurant hygiene grade cards. *The Quarterly Journal of Economics*, 118(2):409–451.

Kane, T. J. and Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *The Journal of Economic Perspectives*, 16(4):pp. 91–114.

Kennickell, A. B. (1998). Multiple imputation in the survey of consumer finances. *Woking Paper*.

Kremer, M. and Holla, A. (2009). Improving education in the developing world: What have we learned from randomized evaluations? *Annual Review of Economics*, 1(1):513–542.

Mas, A. and Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1):112–145.

McEwan, P. J., Urquiola, M., and Vegas, E. (2008). School Choice, Stratification and Information on School Performance: Lessons from Chile. *Economia*.

Meghir, C. and Rivkin, S. (2011). Econometric methods for research in education. volume 3 of *Handbook of the Economics of Education*, pages 1 – 87. Elsevier.

Ministry of Education (2013). Estadísticas de la educación 2013.

Mizala, A. and Urquiola, M. (2013). School markets: The impact of information approximating schools' effectiveness. *Journal of Development Economics*, 103(0):313 – 335.

Neal, D. (2013). The consequences of using one assessment system to pursue two objectives. *Journal of Economic Education*, 44(4):pp. 339–352.

Neal, D. and Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics*, 92(2):263–283.

Neilson, C. (2017). Targeted vouchers, competition among schools, and the academic achievement of poor students. Manuscript.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78(2):311–329.

Quezada-Hofflinger, A. and Von Hippel, P. T. (2018). The response to high stakes testing in Chile, 2005-2013: Legitimate and illegitimate ways to raise test scores. Manuscript.

Rockoff, J. and Turner, L. J. (2010). Short-run impacts of accountability on school quality. *American Economic Journal: Economic Policy*, 2(4):119–147.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley & Sons.

Sánchez, C. (2018). Targeted or universal? mobilizing students through school vouchers. Manuscript.

Sánchez, C. (2019). Skipping your exam? The unexpected response to a targeted voucher policy. *Working Paper*.

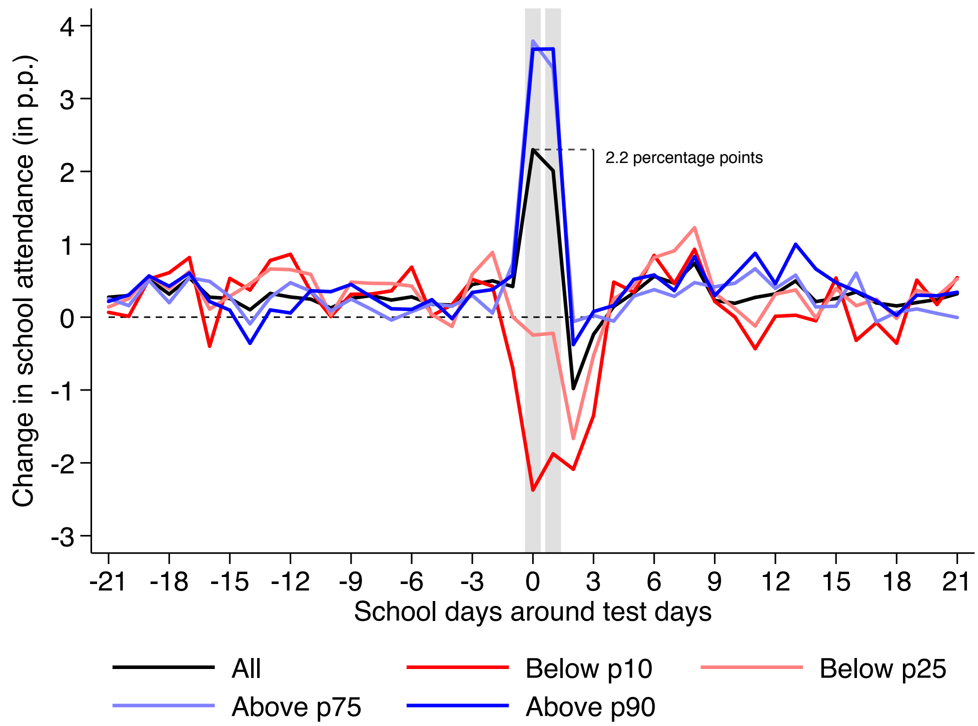
Shleifer, A. (2004). Does competition destroy ethical behavior? *American Economic Review*, 94(2):414–418.

Small, K. and Rosen, H. (1981). Applied welfare economics with discrete choice models. *Econometrica*, 49(1):105–30.

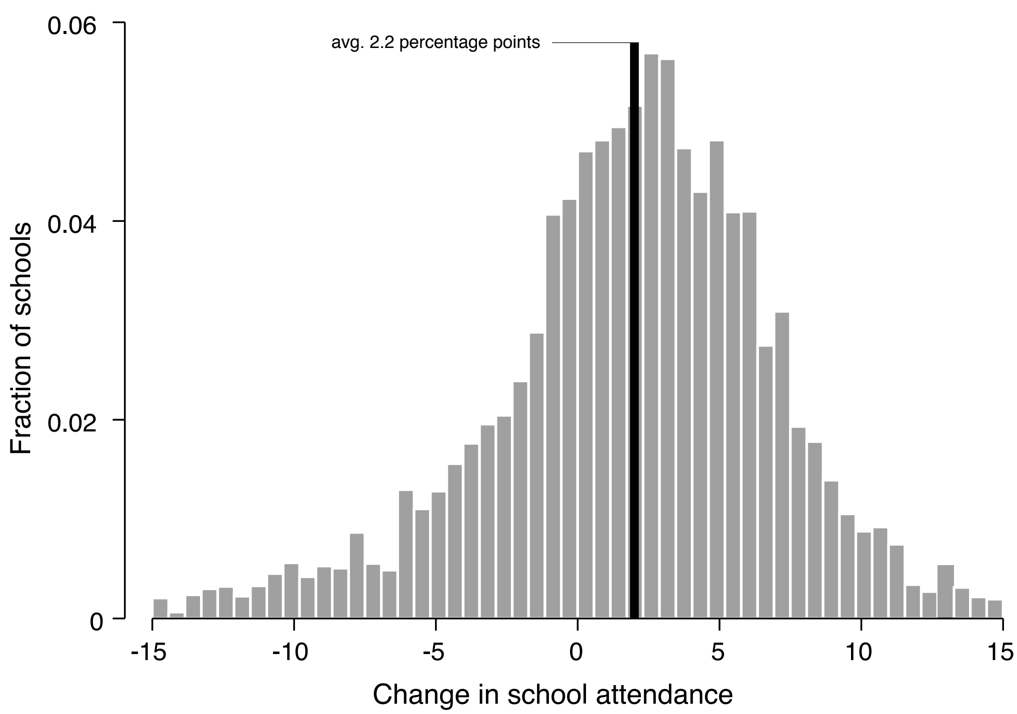
Train, K. (2015). Welfare calculations in discrete choice models when anticipated and experienced attributes differ: A guide with examples. *Journal of Choice Modelling*, 16:15–22.

Walters, C. (2017). The demand for effective charter schools. Manuscript.

Figure 1: School attendance around test days

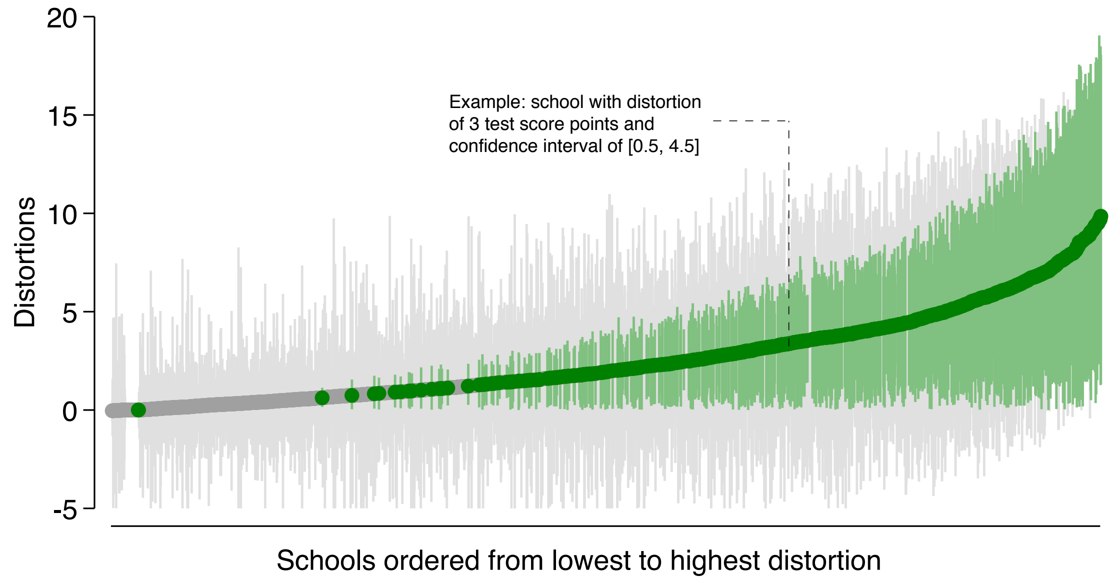


(a) Difference in average attendance rate (y-axis, in percentage points) between 4th graders (test takers) and 3rd graders (non-takers) around the two test days in 2013 (x-axis). Students are grouped by their position in the school GPA distribution, e.g. "Below p10" means students below the 10th percentile of the GPA distribution.

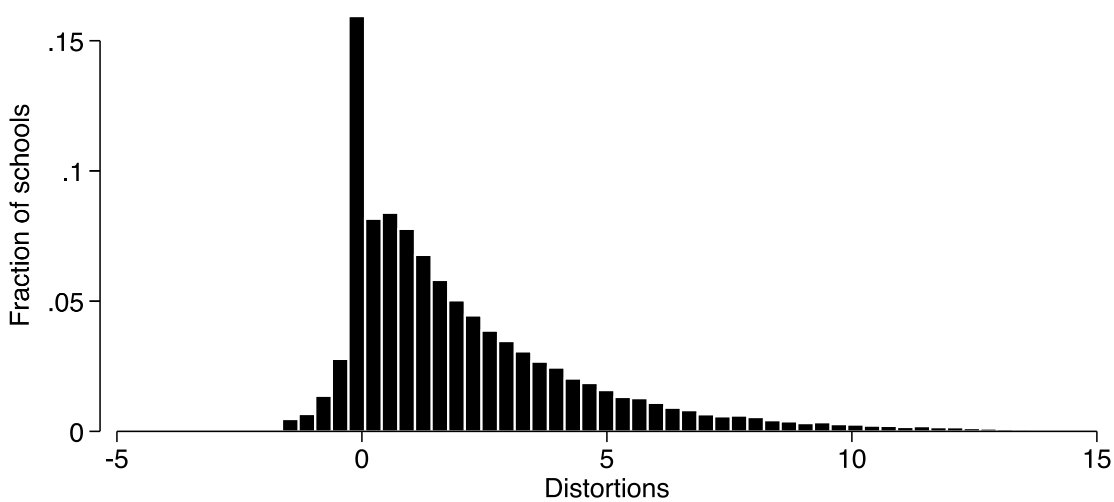


(b) Distribution of changes in school attendance in test days in 2013 (in percentage points).

Figure 2: Distortions in quality signals

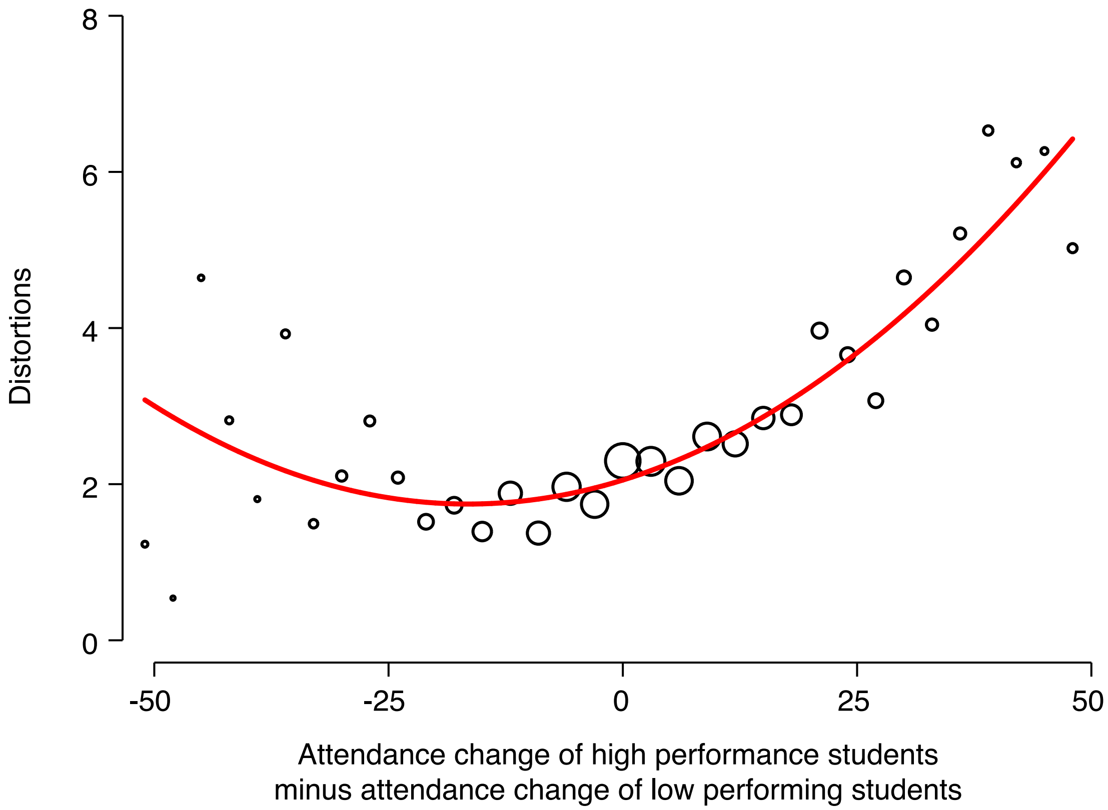


(a) Distortion in quality signals (y-axis, in test score points) are defined as (minus) the difference between school's observed test score and school's counterfactual test score. Schools are ordered from lower to higher distortions in the x-axis. Vertical lines represent the 95 percent confidence interval. Green (gray) lines represents distortions that are (not) statistically different from zero. The figure includes a random sample of distortions for 3,000 school-years.



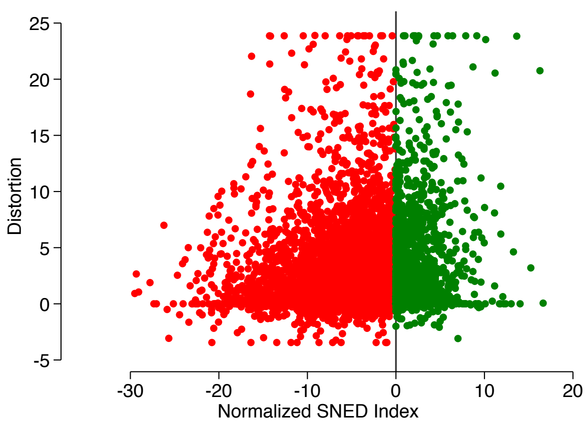
(b) Distribution of distortions in quality signals. Each observation is a school in a specific year between 2005 and 2013.

Figure 3: Distortions and attendance in test days

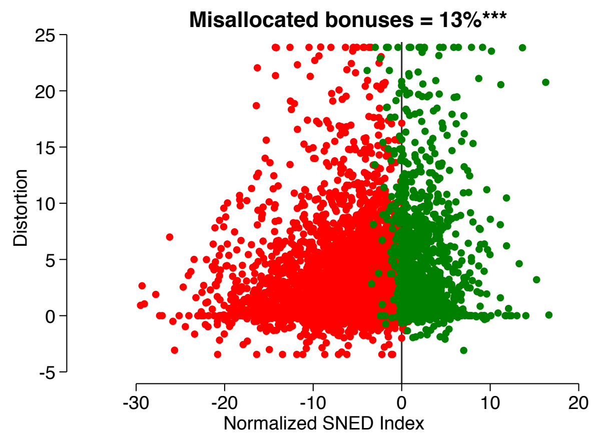


Notes: This figure displays the differential test-day attendance of students above the 75*th* percentile and below the 25*th* percentile of the GPA distribution (x-axis, in percentage points) and distortions in quality signals (y-axis, in test score points). We include all schools in 2013. The coefficients (robust standard errors) of a linear regression of distortions on a linear and quadratic term of differential changes in test-day attendance are 4.38 (0.36) and 3.96 (1.19) respectively. This figure represents a bridge between our test day attendance analysis and distortions and we emphasize we do not use 3rd grade attendance to calculate distortions.

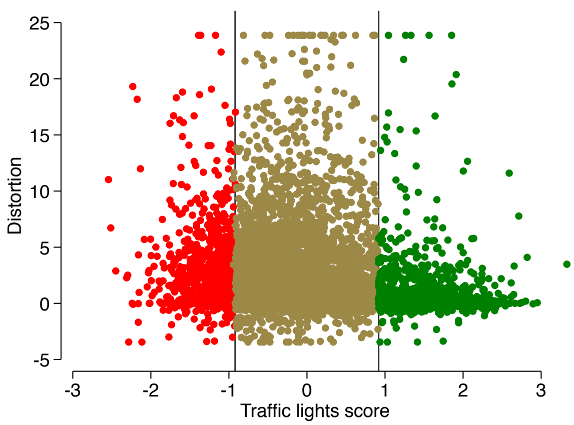
Figure 4: Misallocation of public programs



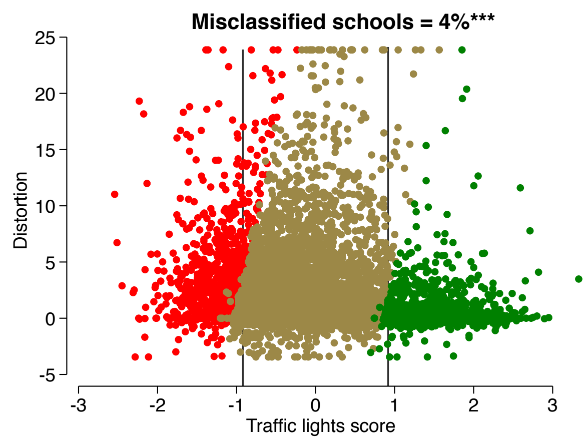
(a) Teacher bonuses (actual assignment)



(b) Teacher bonuses (counterfactual)



(c) Information (actual provision)



(d) Information (counterfactual)

Notes: In panels (a) and (b) we plot school distortions (y-axis), school scores to assign teacher bonuses (x-axis), and the threshold of the assignment (red schools did not get bonuses, green schools did get bonuses) using the actual and counterfactual quality signals. In panels (c) and (d) we plot school distortions (y-axis), school scores (x-axis), and their actual and counterfactual categories (red, yellow, and green).

Table 1: Descriptive statistics

	Observations	Mean	St. dev.	p10	p50	p90
A – Schools (2005-13)						
Test score (SIMCE)	38,416	254.8	27.7	219.5	254.0	292.5
Students in 4*th* grade	38,616	50.4	35.5	17.0	40.0	91.0
Students absent in test days	38,616	3.7	4.5	0.0	3.0	8.0
Class size	38,609	30.4	8.0	19.4	31.0	40.3
Average annual attendance	38,616	93.3	3.1	89.6	93.6	96.7
Students in 1*st*–8*th* grades	38,616	415.5	283.8	143.0	335.0	748.00
Public	38,616	0.39	0.49	0.0	0.0	1.0
Voucher	38,616	0.52	0.50	0.0	1.0	1.0
Private	38,616	0.09	0.28	0.0	0.0	0.0
Religious	37,401	0.44	0.50	0.0	0.0	1.0
Monthly fee (U.S. dollars)	38,341	48.46	92.3	0.0	0.0	182.1
Distortion in test score	60,813	2.7	4.2	0.0	1.1	7.7
B – Students (2013)						
Test score (SIMCE)	140,982	263	46	200	267	321
GPA	159,356	5.9	0.6	5.1	5.9	6.5
Attendance in test-day	137,604	0.95	0.20	1.0	1.0	1.0
Attendance in non-test days	137,127	0.92	0.17	0.8	1.0	1.0

Notes: Own construction based on administrative data provided by the Ministry of Education. We restrict the data to schools with zero distortion or with sufficient data to calculate it. Distortions are measured in test score points and we estimated them using the methodology described in section 4.1. See Figure A.2 for a timeline of standardized tests. See section 4 for details. There are 8,254 schools in the period 2005–2013.

Table 2: Understanding distortions

Dependent variable: distortions in quality signals (in test score points)

	All		Distortions> 0	
A – School attributes	(1)	(2)	(3)	(4)
Public	1.41***	1.26***	0.65**	0.54
	(0.13)	(0.14)	(0.31)	(0.33)
Religious	0.03	-0.11	0.18	0.03
	(0.08)	(0.08)	(0.17)	(0.20)
For-profit	0.35***	0.40***	0.91***	0.94***
	(0.11)	(0.11)	(0.31)	(0.32)
Log parents income	-0.65***	-0.60***	-0.71***	-0.79***
	(0.04)	(0.05)	(0.10)	(0.13)
Average annual attendance	-0.08*	-0.19***	-0.20**	-0.29**
	(0.04)	(0.05)	(0.10)	(0.12)
Students in 4*th* grade	-0.17	-0.11	-2.37***	-2.21***
	(0.15)	(0.15)	(0.25)	(0.26)
Enrollment in grades 1*st*-8*th*	-0.43***	-0.52***	0.04	0.14
	(0.16)	(0.16)	(0.25)	(0.26)
Indicator SEP	1.12***	0.59***	1.42***	0.62*
	(0.07)	(0.10)	(0.15)	(0.35)
Constant	1.22***	1.64***	5.35***	5.97***
	(0.13)	(0.14)	(0.31)	(0.39)
B – Autocorrelation				
Lagged distortion	0.41***	0.38***	0.39***	0.37***
	(0.01)	(0.02)	(0.03)	(0.03)
Constant	1.97***	2.06***	6.25***	6.30***
	(0.04)	(0.05)	(0.13)	(0.14)
Mean of dep. variable	2.18	2.18	5.11	5.11
Market-year F.E.	No	Yes	No	Yes
Variance explained by schools F.E.	0.36	0.36	0.60	0.60
F-test school F.E.	4.59	4.59	2.96	2.96
Schools	3,417	3,417	2,339	2,339
Observations	29,588	29,579	5,929	5,927

Notes: Estimation includes all urban schools. All non-indicator variables have been normalized (except for lagged distortion). All regressions are weighted by the inverse of the uncertainty associated to the calculation of distortions, where uncertainty is the size of the confidence interval. Columns 3-4 restrict the data to school-year observations with distortions statistically different from zero. Robust standard errors in parentheses. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Means of predicted school attributes of households choices

		(1)	(2)	(3)	(4)
		Low-income students		High-income students	
Attribute	Scenario	Switchers	Average	Switchers	Average
Distance	Baseline	2.00	2.36	2.2	2.58
(in kilometers)	Counterfactual	2.07	2.36	2.31	2.59
	Change	0.07	0.00	0.11	0.00
Fee	Baseline	6.58	17.08	39.02	71.43
(in U.S. dollars)	Counterfactual	22.52	17.52	81.95	72.89
	Change	15.95	0.43	42.93	1.47
Quality	Baseline	242.72	254.77	252.62	267.13
(in test score points)	Counterfactual	260.15	255.25	271.53	267.78
	Change	17.43	0.48	18.91	0.65

Notes: Columns 1 and 3 (2 and 4) display the average attributes of chosen schools for low- and high-income switchers (low- and high-income households). Results for distance are measured in kilometers, results for school fees are measured in US dollars and results for quality are measured in SIMCE test scores, net of distortions.

Table 4: Yearly welfare gains of a policy that provides undistorted quality signals

	(1)	(2)	(3)	(4)	(5)	(6)
	Low-income students			High-income students		
Comparison	Switch rate	Switchers $E[\Delta CS_i]$	Average $E[\Delta CS_i]$	Switch rate	Switchers $E[\Delta CS_i]$	Average $E[\Delta CS_i]$
Counterfactual scenario	3.25%	$53.2	$1.73	3.04%	$173.94	$5.29
Low-income households with high-income quality preferences	4.10%	$100.94	$4.13	-	-	-
Low-income households with high-income preferences	3.81%	$181.37	$6.91	-	-	-
Low-income households with high-income market opportunities	2.39%	$ 64.95	$1.55	-	-	-

Notes: Changes in consumer surplus are measured in U.S. dollars per year. Columns 1 and 4 display the share of switchers for low- and high-income households respectively. Columns 3 and 6 display average welfare gains for low- and high-income households. Columns 2 and 5 display average welfare gains for low- and high-income switchers.

For Online Publication

Distorted Quality Signals in School Markets

José Ignacio Cuesta, Felipe González, and Cristián Larroulet

Contents

List of Figures

List of Tables

A More about Estimating Distortions

A.1 Construction of Bounds in Multiple Imputation Method

Let q_{ijt} be the test score of student i in school j and year t. As discussed in the paper, we predict the test score of absent students using the following equation:

$$\widehat{q}_{ijt} = x'_{ijt}\widehat{\gamma}_{jt}$$

where $\widehat{\gamma}_{jt}$ is a vector of parameters estimated by OLS in the sample of students that took the test, and x_{ijt} represents observable variables. To construct bounds for distortions, we take $S = 100$ draws of $\widehat{\gamma}_{jt}$ from the distribution $N(\widehat{\gamma}_{jt}, \widehat{\Sigma}_{jt})$, where $\widehat{\Sigma}_{jt}$ is the estimated variance-covariance matrix for $\widehat{\gamma}_{jt}$. As a result, we have one hundred estimated test scores for each student that did not take the test and, by calculating the average test score for each school-year, one hundred undistorted quality signals. We construct bounds for distortions using the percentiles 2.5 and 97.5 of these one hundred undistorted signals.

A.2 Robustness of Distortions

In this appendix, we develop a variety of exercises in order to assess the robustness of distortions in school test scores, estimated as discussed in Section 4.1.

Model specification for estimation of distortions. Two statistical exercises provide support for the specification in equation (3). First, the R-squared of the 7,500 linear regressions we estimate are high (approximately 0.51) and are always higher in the polynomial rather than the linear model (see Figure A.7). Second, we implement a cross-validation exercise in which we assume test takers are the universe of students and we proceed to delete the test scores of ten percent of students with low GPA, essentially mimicking real world patterns. Reassuringly, in this exercise the quadratic polynomial specification has a lower mean squared error than the linear model. In addition, predicted test scores are similar to observed test scores for students with low and high academic performance, as displayed by Figure A.8.

Accounting for selection into test day. A concern with equation (3) is that of selective attendance. To test for selection, we re-estimated this equation using a Heckman selection correction and found evidence supporting our model. The excluded variable when calculating the Heckman corrected distortions is an indicator for students living outside of the school's county, which effectively predicts attendance on test days. These Heckman corrected distortions are remarkably similar to the uncorrected ones –but noisier, as expected– and both are highly correlated, as displayed by Figure A.11. Finally, our cross validation exercise shows remarkably similar results for both models in terms of mean squared error. Given this evidence, we utilize distortions estimated without this selection correction.

Distortions are independent from noise in test scores. Measurement error (i.e. noise) can also cause discrepancies between observed and true quality signals. However, we emphasize that (1) noise is a mean zero random error that is mean independent of distortions, and (2) distortions are policy-relevant while noise is not. As our setting allows us to calculate the variance of noise in school test scores, we can show the former empirically. We construct a noise distribution for each school in our data using administrative estimates of noise in student-level test scores. These estimates are called "individual-level variability in test scores" and can be aggregated to construct measures of school-level noise following the method in Quality Education Agency (2013). Figure A.12 shows that noise is uncorrelated with distortions (correlation is 0.02), which supports the notion that our analysis of test day attendance represents a different margin that distorts quality signals.

B Understanding Distortions in Quality Signals

B.1 Incentives Placed by SEP Voucher Program

We describe the SEP program in section 2 of the paper. The program generates incentives to raise average test scores through two channels associated with the government funding it provides. In particular, incentives placed by the SEP program operate through a classification of schools that is based largely on SIMCE test scores. This classification then determines (i) the degree of autonomy that schools are provided in spending government funding offered by the program, such that schools with higher test scores have more flexibility than those with lower test scores; and (ii) the renewal of the affiliation of schools to the program after four years in it also depends on SIMCE test scores. For a detailed discussion of the program, see Correa et al. (2014).

The program started in 2008, and while almost all public schools adopted the program immediately, private schools were allowed to choose if and when to adopt it and did so in a staggered fashion across subsequent years. The share of voucher schools in the program increased from 46 percent to 69 percent between 2008 and 2013. We exploit this variation in the timing of adoption of the SEP program by schools to study whether such event has an effect on distortions in quality signals. In particular, we estimate the following event study specification:

$$\psi_{jt} = \sum_{\tau=\underline{\tau}}^{\bar{\tau}} \beta_\tau D_{jt,\tau} + X'_{jt}\theta + \eta_j + \nu_t + \varepsilon_{jt} \tag{1}$$

where $D_{jt,\tau}$ is a dummy that indicates that school j adopted the SEP program τ years before year t, such that β_τ's are the parameters of interest, which measure the effect of SEP adoption τ years after adoption. This specification includes school and year fixed effects, and a vector of time-varying school control variables which includes school enrollment, average attendance rates and number of students taking the SIMCE test.

Figure A.18 presents the results from this analysis. We find that after schools adopt the SEP program, they increase their distortions in quality signals by around 0.7 points (0.17σ). This increase persists four years after adopting the program. These results suggest that pressures associated with the SEP program may induce schools to introduce distortions in their test scores in order to comply with requirements set by the SEP program for renewing affiliation with it and continue receiving additional government funding.[1] We interpret these results with caution, given that adoption of the SEP program is a choice of schools (in particular, for voucher schools) and therefore program adoption could be correlated with school unobservables also driving distortions. However, the pre-trend leading to adoption is remarkably well behaved, which limits that concern. These results are consistent with recent work by Feigenberg et al. (2018), Quezada-Hofflinger and Von Hippel (2018) and Sánchez (2019).

B.2 Monetary Incentives for Teachers

We describe the SNED program in section 2 of the paper. Given that (i) prizes are provided according to an index, and (ii) after each contest schools are informed of their outcomes, we can use a school's index as a measure of incentives. We compute the distance of each school to the threshold for obtaining the prize. Schools closer to the threshold have more incentives to increase their test scores through distortions than those further away from the threshold either upwards (sure winners) or downwards (sure losers). Using this rationale, we estimate:

$$\psi_{jt} = \mathbf{1_{IN}} f^{IN}(\text{SNED}^{IN}_{jt-1}) + \mathbf{1_{OUT}} f^{OUT}(\text{SNED}^{OUT}_{jt-1}) + \eta_j + \nu_t + \varepsilon_{jt} \tag{2}$$

where SNED^{IN}_{jt-1} measures distance to the threshold for winners, and SNED^{OUT}_{jt-1} measures distance to threshold for losers, both in terms of index points. We use information from the previous contest to construct these variables. Our objects of interest are the functions f^{IN} and f^{OUT}. If schools closer to the threshold have larger distortions, we would interpret it as evidence of teachers introducing distortions to test scores as a response to the incentives placed by the program.

Figure A.19 presents four different plots for the relationship between distortions and schools' distance to the threshold. We present results for the two years after the prize is awarded and both for raw distortions in quality signals and residualized distortions (net of school and year fixed effects, as well as school characteristics). Estimates of f^{IN} and f^{OUT} show, if anything, the opposite pattern: schools closer to the cutoff have lower or similar distortions to quality signals. These results provide suggestive evidence against the hypothesis that teachers manipulate attendance to increase test scores.

B.3 Information for School Choice

Other quality disclosure policies could incentivize schools to introduce distortions in quality signals, as is the case of the ETL informational policy, which we use to test for this mechanism. See section 2 of the paper for details about the program.

[1] The implementation of the SEP program was not as rigorous as intended. In fact, the first sanctions to schools were implemented in 2012, several years after the program started (Neilson, 2017b). These results suggest that, regardless of difficulties with the implementation, schools at least partially believed that MINEDUC would behave in accordance to the design of the program according to sanction, as in absence of such belief there would have been no additional pressure on schools regarding increasing test scores and, therefore, no increase in distortions in quality signals.

Following the discontinuous incentives at the threshold, we estimate:

$$\psi_{jt} = \mathbf{1_r} f^r(q_{jt-1}) + \mathbf{1_y} f^y(q_{jt-1}) + \mathbf{1_g} f^g(q_{jt-1}) + X'_{jt}\theta + \varepsilon_{jt} \tag{3}$$

where q_{jt-1} measures test scores with which the ETL policy was assigned to schools. Our objects of interest are the functions f^r, f^y and f^g, where r, y and g stand for the three different quality levels signed by the policy to schools. If schools closer to the policy thresholds have larger distortions, we would interpret it as evidence of schools introducing distortions in order to signal a higher level of quality in a subsequent version of the policy.

Figure A.20 presents the linear relationships between test scores and distortions around the ETL policy cutoffs. Again, we present results for distortions and residualized distortions.[2] These plots show that distortions increased slightly around the cutoff between red and yellow schools. This means that schools introduce larger distortions in order to move towards the yellow category or avoid moving to the red category. Note that once school characteristics are controlled for, this pattern can hardly be noticed. This pattern, however, is not the same around the second cutoff. These results do not provide strong evidence that schools closer to thresholds set by this policy introduce higher distortions in order to signal higher quality.

C School Choice Model

We develop estimate a model of school choice in the lines of Bayer et al. (2007), which we then exploit in Section 5 to study a counterfactual in which we provide undistorted information to households and analyze how that change choices and welfare. When constructing the model, we impose certain assumptions, some of which are related to the Chilean institutional framework. First, we assume that households are informed regarding both available schools and their *observed* characteristics. Distortions or information to infer them are unobserved by households. Second, we assume that schools do not select students based on attributes and do not face capacity constraints, i.e. households can enroll their children in any school in their choice set. As discussed by Gallego and Hernando (2009) and Neilson (2017a), this assumption is likely to hold in the Chilean school system. Finally, we assume the household's location choice is independent of the school choice problem. This assumption is supported by the lack of constraints on the choice set of schools based on residential location.

Let households be indexed by i and schools by j. Household utility depends on school fees, quality, and distance to school, denoted respectively p_j, q_j and d_{ij}. They also derive utility from other school characteristics W_j. For notational simplicity, we denote $X_j = [p_j, q_j, W_j]$, which includes K attributes. Preferences are heterogeneous depending on household type, indexed by r. In our model, only observed heterogeneity in preferences is considered, as explained below. Moreover, we allow for households to derive utility from schools' characteristics that are unobserved to the econometrician, ξ_j. Finally, each household has an idiosyncratic preference shock, ε_{ij}, which we assume is distributed Type-1 Extreme

[2] Note that this is a cross-sectional exercise, so we cannot include school and year fixed effects in this case, just school characteristics.

Value (T1EV).

Under these assumptions, the indirect utility of household i of type r from enrolling their children in school j is:

$$u_{ij}^r = \sum_k x_{k,j}\beta_k^r + \xi_j^r + \beta_d^r d_{ij} + \varepsilon_{ij} \tag{4}$$

where the first two terms measure utility from characteristics that depend only on the school and are therefore constant across households of type r for a given school j, while the third term measures disutility from distance between household i and school j for households of type r, which varies across households. We can therefore rewrite equation (4) as follows:

$$u_{ij}^r = \delta_j^r + \beta_d^r d_{ij} + \varepsilon_{ij} \tag{5}$$

such that the parameters of the model are contained in the vector β^r, but can be alternatively represented by the vector δ^r and by β_d^r. Note that δ_j^r is the component of utility derived from choosing school j that is constant across households, the mean value of school j for households of type r.

The probability of household i choosing school j can be derived analytically using households indirect utility.[3] The choice probability of school j by household i of type r predicted by the model is a function of school and household characteristics:

$$P_{ij}^r(\delta^r, \mathbf{d^r}, \beta_d^r) = \frac{\exp(\delta_j^r + \beta_d^r d_{ij})}{\sum_{l\in\mathcal{J}_i}\exp(\delta_l^r + \beta_d^r d_{il})} \tag{6}$$

where $\mathcal{J}_i$ is the set of schools in the market where household i is located. We use this result in the next subsections for both estimating the model and for computing the counterfactual exercise of interest.

C.1 Estimation

We estimate the parameters of the model using a two-step procedure. First, we estimate standard conditional logit models for each group r in each market and year in the data, to recover schools' mean values. Second, we exploit the assumed linear functional form of households' indirect utility function in order to estimate the relationship between schools' mean values and and school attributes and recover preference parameters .

The first stage of the estimation procedure consists of estimating equation (6), which can be done by maximum likelihood. In order to allow for heterogeneity in preferences, this procedure is implemented within each of multiple cells defined on the basis of R socioeconomic levels, T time periods, and M markets. The former is determined by the eligibility of a student for the SEP program, which is determined by participation in social programs aimed at supporting low-income households. Therefore, we estimate $R \times T \times M$ conditional logit models in the first stage, which yields the same number of

[3]In the context of school choice, there is no obvious outside option. Therefore, we instead normalize $\delta_1 = 0$ within each market.

estimates for δ^r and β^r_d.

The second stage exploits the assumed linear functional form of the utility function in order to estimate the following linear regression:

$$\delta^r_{jmt} \quad = \quad \delta^r_{0,mt} + \sum_k x_{k,jmt}\beta^r_k + \epsilon^r_{jmt} \tag{7}$$

where $\delta^r_{0,mt}$ is a constant term specific to each market, year, and household type; β^r_k measures the effect of x_k on school mean value for households of type r and maps to the preference parameters of our model; and ϵ^r_{jmt} is a mean-zero error term. Note that $\delta^r_{0,mt} + \epsilon^r_{jmt}$ maps to the unobserved school characteristic ξ^r_{jmt} in our model.

A concern with this type of regression is the potential endogeneity of school characteristics, particularly of prices and quality. Therefore, we estimate this regression using an instrumental variables approach, using various instruments. First, for each school, we include the fixed non-price and non-quality characteristics of other schools in the market, in line with instruments suggested in Berry et al. (1995). In particular, we compute the share of religious schools, schools with gender constraints, and public schools in the market for each school in the sample, using them as instruments. Second, we follow Neilson (2017a) and use average teacher hourly wages, which arguably operates as a cost shifter for schools, such that it might affect their choices of fees. Third, we use the amounts of funding provided by different voucher program components, which display within market variation due to school characteristics that are fixed in the short run. In particular, we include the baseline voucher and two additional components related to a school being part of the SEP program and to a school having a concentration of SEP students above a threshold. Moreover, we utilize county temperature data on test days as an instrument for quality. While the data provides support for a relationship between temperature and test scores, it would be hard to argue that temperature on test days could otherwise be correlated with unobserved school attributes. This instrument is motivated by a literature that studies the relationship between climate and academic achievement (Graff Zivin et al., 2018; Park, 2018).[4] Finally, we use an indicator variable for whether a school was awarded a SNED prize in its most recent version. This instrument is motivated by Contreras and Rau (2012) who show how these prizes impact quality in subsequent years.[5]

We estimate the model using data for 2011 through 2014, the only years in which student home address data is available. In addition, we only utilize data for students in $1st$ grade in order to focus on the margin in which most school choices are made. In terms of covariates to be included in the vector X_j, we include school fees, quality as measured by the school's average SIMCE test score, whether the school has a religious orientation, whether the school has any gender constraints, whether a school is

[4]We construct this variable using data from the Berkeley Earth dataset, which provides population-weighted estimates of daily temperature at the county level. In implementing this regression, we include both temperature and temperature squared in order to account for non-linear effects of temperature on academic achievement as documented in Graff Zivin et al. (2018).

[5]In practice, we utilize the residual of a regression of the SNED award indicator on quality in the year of the award in order to further control for quality differences between SNED awardees and non-awardees which might be driven by other factors that could be persistent in time.

public, and whether a school is part of the SEP program.[6] Finally, we are able to compute the distance between households and schools using geo-referenced data on their addresses.[7]

C.2 Market Definition and Estimating Dataset

Determining which suppliers belong to the consumers' choice set in context of spatial competition is not straightforward. In contrast to other school systems, in Chile there are not any institutional constraints that limit the extent to which students can travel. Therefore, we need to define markets.

We adopt an approach based on the spatial distance between schools, similar to that in Neilson (2017a). Distance has been shown to be a relevant determinant of school choice in the literature (Gallego and Hernando, 2009; Neilson, 2017a). In our data, students' average distance to chosen schools is 1.3 miles and the 90*th* percentile of such distribution is 3 miles. Therefore, it makes sense to argue that schools located far enough from each other might belong to different educational markets. We define an educational market as a cluster of schools in a closed polygon with no other school closer than 3 miles from its boundaries. Operationally, a market is uniquely identified from the adjacency matrix of schools, where links are defined as two schools being closer than 3 miles from each other. In implementing this procedure, and therefore in estimation as well, we only consider urban schools. Specifically, we only include markets with at least 20 schools and for which we have data for at least 300 students. The map presented in Figure A.21 provides an example for the resulting market definitions, and Table A.2 displays its summary statistics.[8]

A description of the resulting sample is displayed in Table A.3. The number of household types is $R = 2$, the number of markets included is $M = 25$, and the number of periods covered is $T = 4$. Therefore, the estimating dataset is comprised of 200 cells. The estimating dataset includes 1,556 schools and 97,471 students. On average, 33 percent of the students attending schools in markets in our sample are included, and 92 percent of the schools operating in each market. Moreover, an average of 49 percent of students included in the sample across markets are eligible for the SEP program.[9]

[6]We use data on monthly copayments faced by households as a measure of school fees. Moreover, we use data on students' eligibility for SEP in order to adjust school fees accordingly; eligible students do not pay any school fees in schools that operate under the SEP regime.

[7]We compute the Euclidean distance between every household and school in each market. We then proceed to clean these results by (i) removing mass points, which arise from imperfect geo-reference; and (ii) removing students located further than 55 kilometers from the median household location in the market.

[8]As a robustness exercise, we estimated the model using counties as markets. For estimation, we included counties for which a large share of students resided in the market (at least 90 percent) and where we had available data for more than 300 students. Results were quantitatively similar.

[9]We tested for differences in observables across students included and excluded in the sample within each market. While some of the differences across groups are statistically significant, they are not economically significant and do not show a clear pattern. Results are available upon request.

C.3 Results

Given that the most relevant dimension of household heterogeneity is socioeconomic status, we present all the results for low- and high-income households separately. Figure A.22 displays the resulting coefficients in each market for distance between households and schools for both low- and high-income households. In all these cases, the coefficient is negative, which reflects a decreasing utility for choosing a school further away from home. Low-income households are on average 14 percent more distance-sensitive than high-income households.

Table A.4 presents results for different specifications of instrumental variables linear regressions of the estimates of δ_{jmt} on different sets of school characteristics and fixed effects. Columns 1 through 3 display results for all households in the sample, columns 4 through 6 display results for low-income households, and columns 6 through 9 for high-income households. Overall, results point in the expected direction: household utility decreases with school fees and increases with their reported quality. Both adding market-year fixed effects and other school attributes to the regression increase the magnitude of point estimates with respect to the baseline case.[10] Overall, the model provides a good fit for observed enrollment shares, as displayed by Figure A.23. The correlation between observed and predicted enrollment shares is of 0.88.

There are interesting patterns of heterogeneity across low- and high-income households. For example, our preferred specifications in columns 6 and 9 imply that low-income households are 88 percent more price-sensitive than high-income households. Inversely, low-income households are estimated to be 37 percent less quality-sensitive than high-income households. These results imply in turn that high-income households' willingness to pay for quality is three times higher than that of low-income households. This heterogeneity suggests that quality disclosure policies will have heterogeneous effects across these demographic groups. These patterns of heterogeneity coincide with previous findings within the school choice literature (e.g. Gallego and Hernando 2009, Hastings et al. 2009, and Neilson 2017a).[11]

C.4 Details on Counterfactual

We compute effects on school choices by adjusting the choice probabilities predicted by equation (6) our model and using parameter estimates and data on school attributes for both scenarios. Following equation (6), choice probabilities are therefore computed as $P^r_{ijmt}(d^r, \hat{\delta}^r, \hat{\beta}^r_d)$ and $P^r_{ijmt}(d^r, \tilde{\delta}^r, \hat{\beta}^r_d)$, where $\tilde{\delta}^r_{jmt} = \sum_k \tilde{x}_{k,jmt}\hat{\beta}^r_k + \hat{\xi}^r_{jmt}$ is the mean utility of school j in market m in period t, computed using preferences

[10] Table A.5 and Table A.6 display results from the first stage of the IV estimation for school fees and quality respectively. The bottom rows in Table A.4 show the respective F-tests for the subsets of instrumental variables utilized for school fees and quality respectively. Moreover, we further assess the strength of the instruments by reporting the Cragg and Donald (1993) eigenvalue statistic for each specification. Stock and Yogo (2005) provide critical values for rejection of this test. In our setting, the critical value for rejection is 29.32, always below the reported values for the Cragg-Donald statistic. Finally, Table A.7 displays the results from estimating the second stage of the model by OLS. As expected, the OLS estimates are smaller than the IV ones.

[11] As a robustness check on the results, we study the correlation in estimates of unobserved school characteristics ξ^r_{jmt} across low- and high-income households. While there is variation in results across both groups, there is a positive correlation of 0.57 between the estimates of ξ^r_{jmt}. This is, while services provided by schools might be differently valued across consumer types, those values are strongly correlated across them.

estimates and data on counterfactual school quality. To compute the share of *switchers* in the population, we simulate choices of consumers in our sample in both the baseline and counterfactual scenarios. Reported results correspond to average switching rates for low- and high-income households over 200 simulations across all households in the sample.

References

Bayer, P., Ferreira, F., and McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, 115(4):588–638.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

CIPER (2012). Ley SEP: Contraloría revela masiva pérdida de recursos destinados a educación de los más pobres. *Centro de Investigación Periodística*.

Contreras, D. and Rau, T. (2012). Tournament incentives for teachers: evidence from a scaled-up intervention in Chile. *Economic Development and Cultural Change*, 61(1):219–246.

Correa, J. A., Parro, F., and Reyes, L. (2014). The effects of vouchers on school results: Evidence from Chile's targeted voucher program. *Journal of Human Capital*, 8(4):351–398.

Cragg, J. G. and Donald, S. G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, 9:222–240.

Feigenberg, B., Rivkin, S., and Yan, R. (2018). Illusory gains from Chile's targeted school voucher experiment. Working Paper 23178, National Bureau of Economic Research.

Gallego, F. and Hernando, A. (2009). School choice in chile: Looking at the demand side. *Pontificia Universidad Catolica de Chile Documento de Trabajo*, (356).

Graff Zivin, J., Hsiang, S., and Neidell, M. (2018). Temperature and Human Capital in the Short and Long Run. *Journal of the Association of Environmental and Resource Economists*, 5(1):77–105.

Hastings, J., Kane, T. J., and Staiger, D. O. (2009). Heterogeneous preferences and the efficacy of public school choice. Manuscript.

Neilson, C. (2017a). Targeted vouchers, competition among schools, and the academic achievement of poor students. Manuscript.

Neilson, C. (2017b). Targeted vouchers, competition among schools, and the academic achievement of poor students – online appendix. Manuscript.

Park, J. (2018). Hot temperature and high stakes cognitive assessments. *Working Paper*.

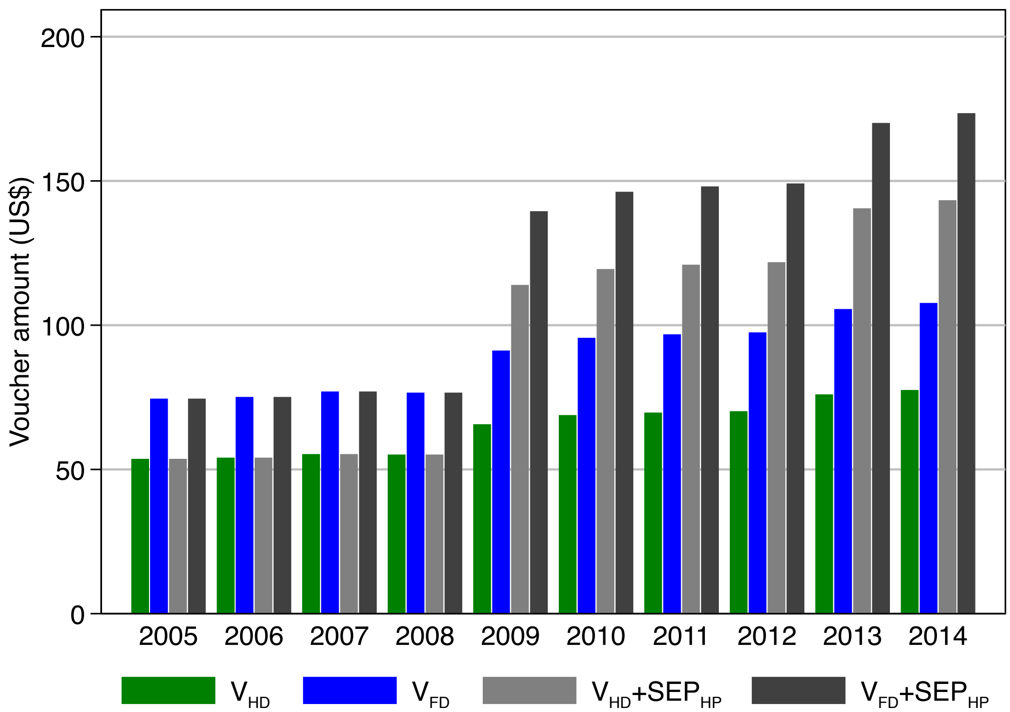
Quality Education Agency (2013). Cálculo de significancia estadística para resultados de las pruebas Simce. *Unidad de Análisis Estadístico, División de Evaluación de Logros y Aprendizaje*.

Quezada-Hofflinger, A. and Von Hippel, P. T. (2018). The response to high stakes testing in Chile, 2005-2013: Legitimate and illegitimate ways to raise test scores. Manuscript.

Sánchez, C. (2019). Skipping your exam? The unexpected response to a targeted voucher policy. *Working Paper*.

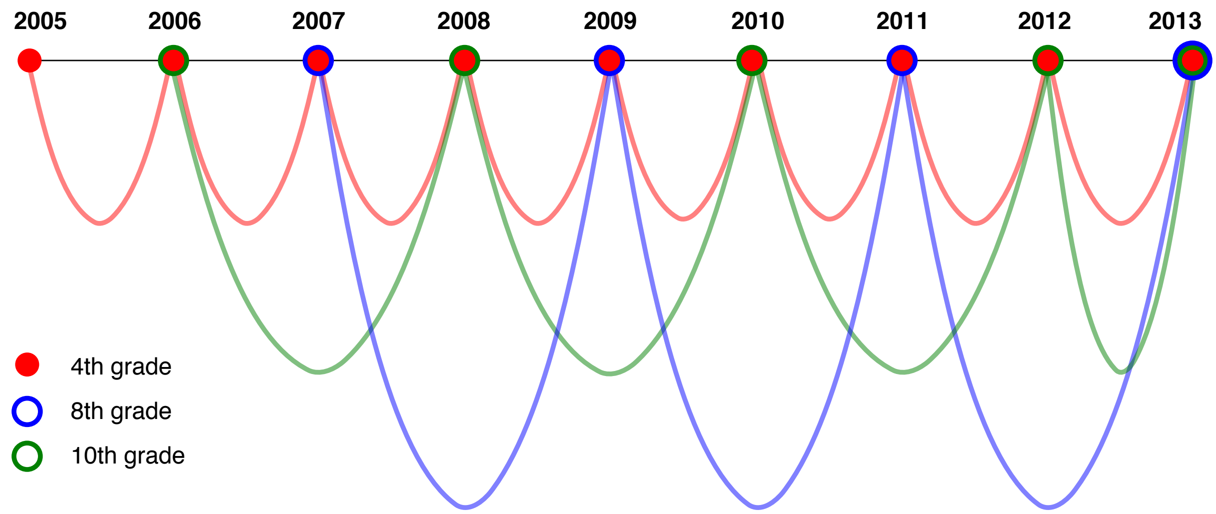
Stock, J. and Yogo, M. (2005). Testing for weak instruments in linear IV regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, chapter 5, pages 80–108.

Figure A.1: Evolution of vouchers



Notes: Amount covered by different types of vouchers in the system. In particular, four types are displayed, covering the interaction of schools offering half and full school shifts (i.e. HD and FD) according to the JEC program, and school subscribed and not subscribed to the SEP program. This figure displays the voucher amount for SEP school with high performance. Note that this figure do not display all voucher types: the voucher amount for low performing SEP schools and the component of SEP vouchers related to the concentration of SEP students in schools are not reported.

Figure A.2: Timeline of standardized test scores



Notes: Year and grade of students taking the national standardized test (SIMCE) in the period 2005–2013. Math and language tests are always taken by students. Natural and social sciences tests are taken by subsets of students. Additional tests have been applied to 2*nd* and 6*th* grade students since 2012, but we omit them from our analysis because they are relatively new.

Figure A.3: Test scores as quality signals

(a) Public dissemination of test scores

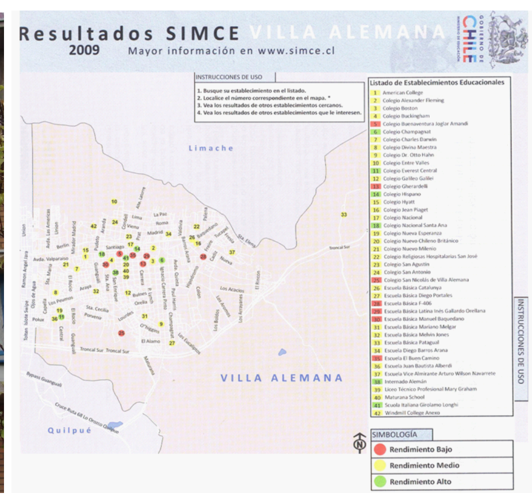
4º Básico Los 50 colegios con mejor Simce

	Establecimiento	Comuna	Dep.	Lenguaje	Matemática	C. Medio Natural	Prom. leng. y matemat.
1	Colegio La Girouette	Las Condes	PP	336	328	332	332,0
2	Colegio San Jorge	Laja	PP	316	345	328	330,5
3	Escuela Básica Madrigal	La Reina	PP	329	329	322	329,0
4	Colegio Curimon	San Felipe	PP	335	322	330	328,5
5	Colegio The Grange School	La Reina	PP	326	324	322	325,0
5	Colegio Kendal English School	Ñuñoa	PP	325	325	316	325,0
7	Colegio Redland School	Las Condes	PP	329	320	307	324,5
7	Colegio Cambridge College	Providencia	PP	327	322	327	324,5
7	Colegio Santiago College	Providencia	PP	331	318	313	324,5
10	Colegio Arauco	Arauco	PP	317	331	322	324,0
10	Colegio Constitución	Constitución	PP	314	334	315	324,0
12	Colegio Itahue de Chiguayante	Chiguayante	PP	317	329	314	323,0
13	Escuela Básica, Colegio Peñalolén	Peñalolén	PP	325	320	314	322,5
13	Colegio Montemar	Viña del Mar	PP	323	322	317	322,5
13	Colegio San Joaquín	La Serena	PP	317	328	316	322,5
16	Colegio Pucón Nevados	Pucón	PS	319	325	307	322,0
17	Souther Oxford School	Villarrica	PP	313	330	316	321,5
18	Colegio Internacional Alba	Maipú	PP	321	321	319	321,0
18	Colegio de La Salle	La Reina	PP	321	321	316	321,0
18	Colegio Arrayanes	San Fernando	PP	323	319	324	321,0
21	Colegio Mozart Schule	Concepción	PP	322	319	301	320,5
22	Colegio Cumbres	Las Condes	PP	317	322	311	319,5
23	Colegio Villa María Academy	Las Condes	PP	320	318	296	319,0
24	Colegio P. Hurtado y Juanita de los Andes	Las Condes	PP	320	317	308	318,5
24	Colegio San Benito	Vitacura	PP	315	322	288	318,5
26	Colegio Andrée English School	La Reina	PP	315	321	330	318,0
26	Inst. de Humanidades Alfredo Silva S.	Concepción	PP	319	317	314	318,0
26	Colegio Alborada	Linares	PP	318	318	304	318,0
26	Saint Paul's School	Viña del Mar	PP	321	315	302	318,0
30	Colegio Buin	Buin	PP	317	317	318	317,0
30	Escuela Básica Trupam	Conchalí	PS	313	321	301	317,0
32	Colegio Akros S.A.	Ñuñoa	PP	319	314	305	316,5
32	The Wessex School	Concepción	PP	325	308	306	316,5
34	Colegio los Andes de Vitacura	Vitacura	PP	319	312	310	315,5
34	Complejo Educacional San Crescente	Castro	PS	305	326	282	315,5
34	Cerro Estanque	Tomé	MD	319	312	296	315,5
34	Colegio Quimahue	Rancagua	PP	311	320	315	315,5
38	Colegio San Ignacio	Calama	PP	299	331	321	315,0
38	Colegio Saint George's College	Vitacura	PP	317	313	302	315,0
38	Colegio Eduardo Martín Abejon	Constitución	PS	331	299	281	315,0
41	Colegio The Antofagasta British School	Antofagasta	PP	315	314	312	314,5
42	Colegio Saint Gabriel's	Providencia	PP	317	311	302	314,0
43	Colegio Pedro de Valdivia Las Condes	Las Condes	PP	313	314	307	313,5
43	Colegio Wenlock School	Las Condes	PP	316	311	286	313,5
43	Colegio Particular Montessori	Talca	PP	318	309	306	313,5
43	Colegio de los Sagrados Corazones	Valparaíso	PP	312	315	312	313,5
47	Colegio del Sagrado Corazón de Apoquindo	Las Condes	PP	314	312	304	313,0
47	Instituto Miguel León Prado	San Miguel	PS	315	311	315	313,0
47	Colegio Vedruna	San Felipe	PS	320	306	307	313,0
47	Scuola Italiana Arturo Dell'Oro	Valparaíso	PP	315	311	313	313,0

Fuente Simce

(b) Test scores in newspaper

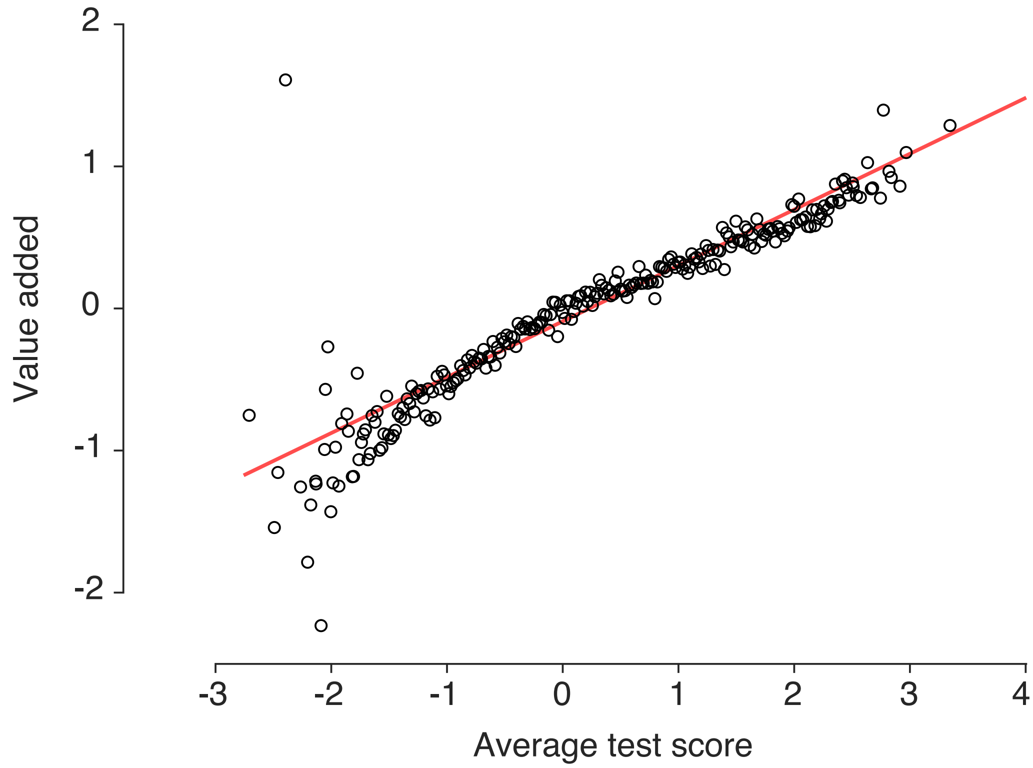
(c) Test scores as an advertising device



(d) Test scores as a policy tool

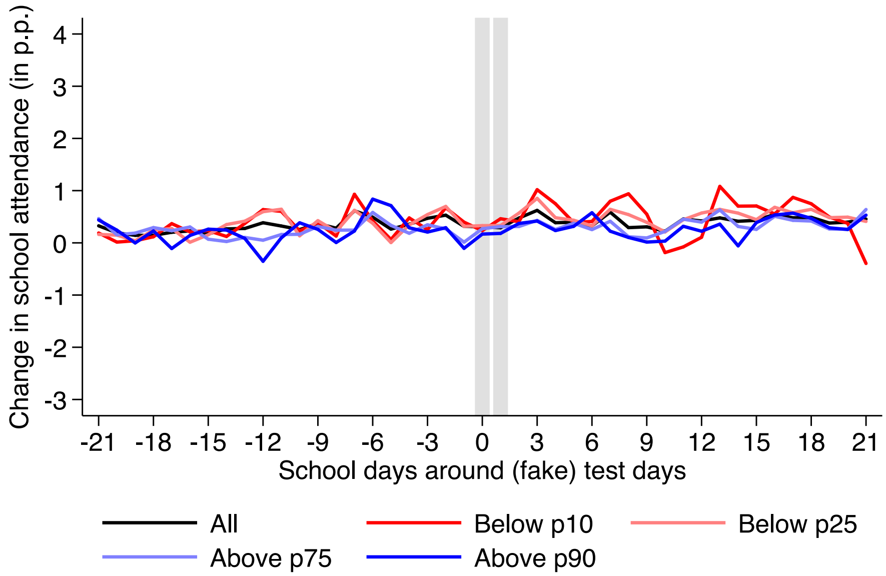
Notes: This figure displays the different roles of test scores in the Chilean educational system. Panel (a) displays the front page of *La Segunda*, a popular newspaper, advertising the disclosure of school level test scores for all schools. Panel (b) shows schools' test scores as published in newspaper *El Mercurio*. Although test scores are observable, other variables such teacher wages, teacher quality, value added, and school composition, are not. Panel (c) displays an advertising banner placed on the front of a school reporting on successful results obtained by the schools in SIMCE as a means of advertising its quality to households. Panel (d) displays an example of of the Educational Traffic Lights policy, which utilizes SIMCE test scores as an input for quality disclosure.

Figure A.4: Correlation between test scores and value added

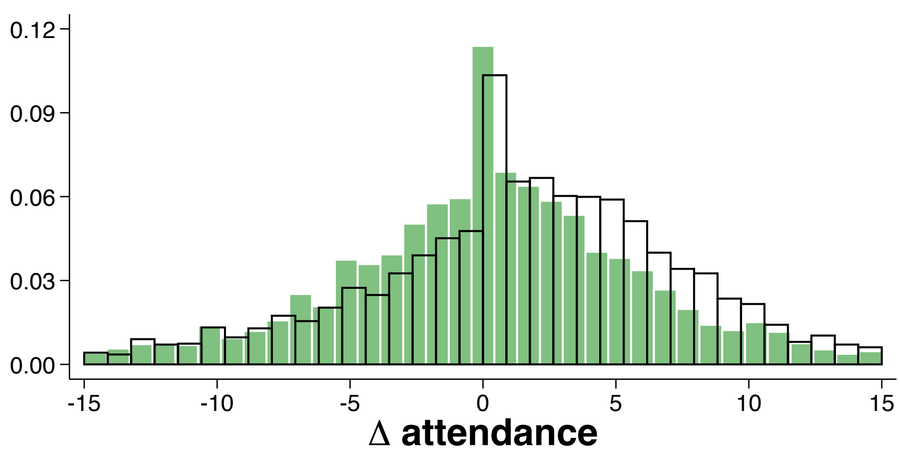


Notes: This figure displays the relationship between test scores and the only available measure of value added in Chile, from Neilson (2017a). We thank the author for providing us with this figure.

Figure A.5: Comparison of school absenteeism on test day with fake test days



(a) Differences-in-differences using school days around April 23th, 2013



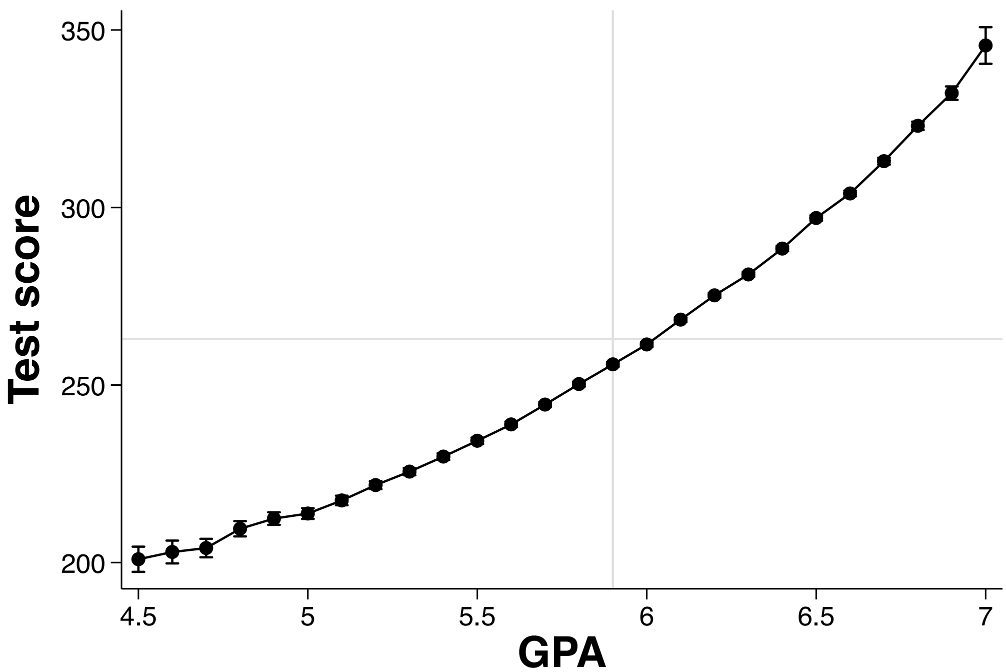
(b) Comparison of distributions (green is April 23th and white are actual test days)

Notes: Panel (a) presents the difference in absenteeism rates between 4*th* and 3*rd* graders across the GPA distribution around April 23th of 2013, a day without standardized tests. The histograms in panel (b) represent the distribution of the following differences-in-differences estimate at the school level:

$$\Delta\overline{A}_j = (\overline{A}_{j4T} - \overline{A}_{j4t}) - (\overline{A}_{j3T} - \overline{A}_{j3t})$$

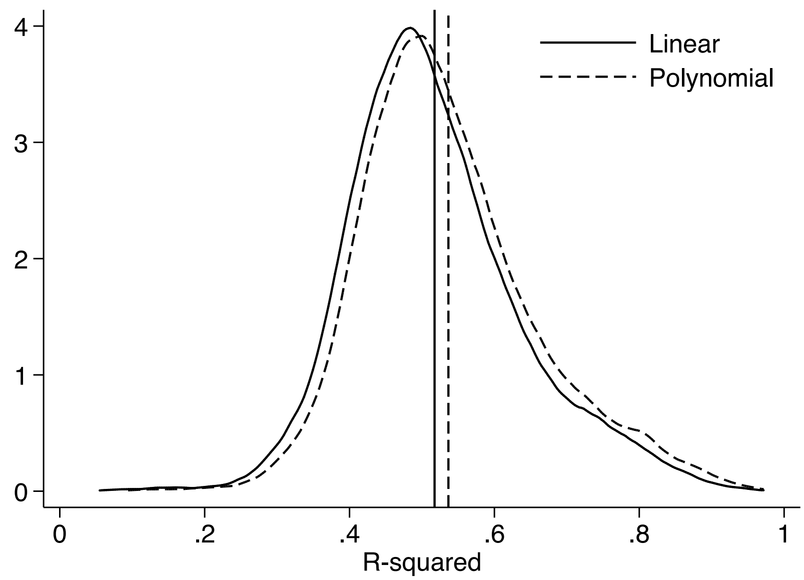
where $\overline{A}_{jkt}$ is the average absenteeism rate of k*th* graders in school j in day t. Day $t = T$ represents the day of the event analyzed (green is April 23th and white actual test days). A Kolmogorov-Smirnov test rejects the equality of distributions in both cases (p-values < 0.01).

Figure A.6: Predictability of test scores

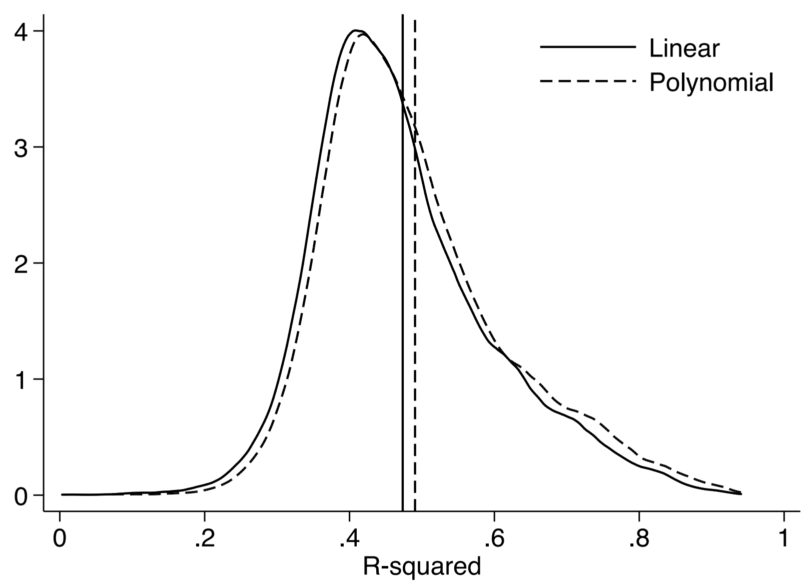


Notes: Coefficient estimates and 95 percent confidence interval of a linear regression of test score on (1) a full set of indicators for a student's GPA, and (2) school fixed effects. Standard errors are clustered at the school level. Gray lines indicate the mean.

Figure A.7: Prediction model



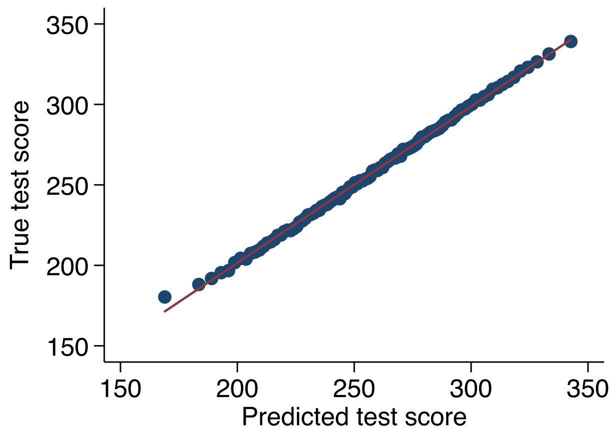
(a) Math



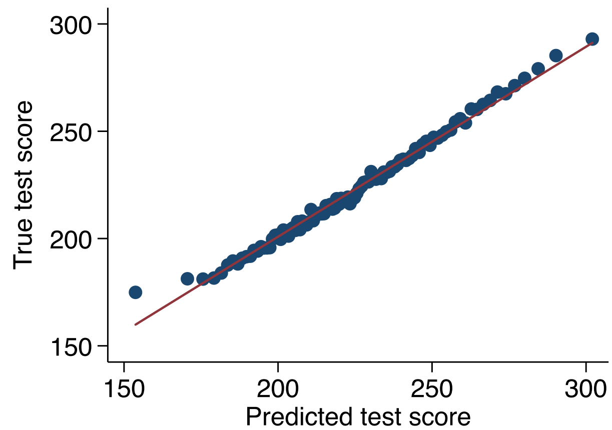
(b) Language

Notes: This figure presents the distribution of R-squared for all regressions of test scores on observable variables (i.e. predictors) among test takers in each school in our data. We include predictors linearly (solid line) or as a polynomial (dash line). Recall that these predictions include GPA, indicators for school switchers and students who are repeating the grade, and year fixed effects. Vertical lines denote the average R-square in the corresponding panel. Panel (a) plots the R-squared for the mathematics test and panel (b) plots the R-squared for the language test. There are a total of 7,493 regressions in each panel.

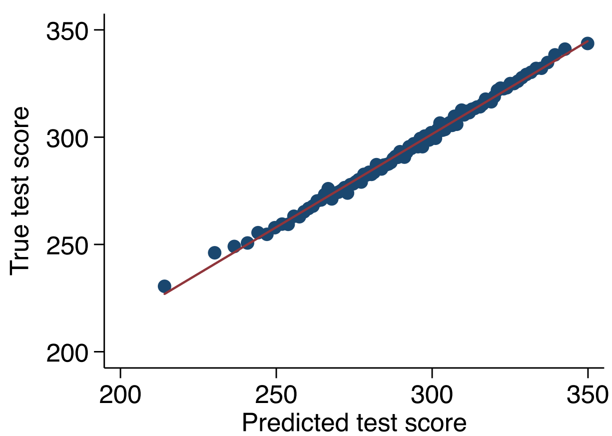
Figure A.8: Evaluation of prediction model



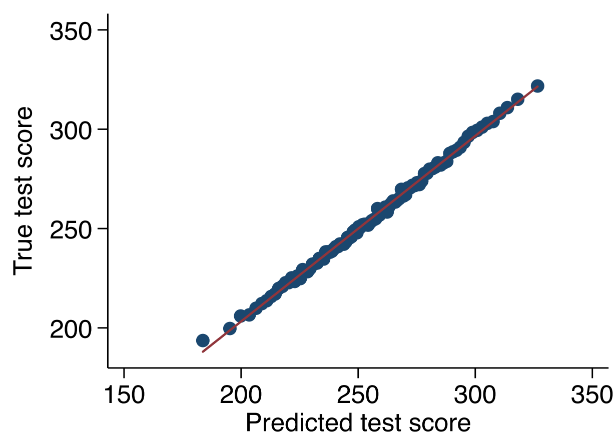
(a) All students



(b) Students with low academic performance



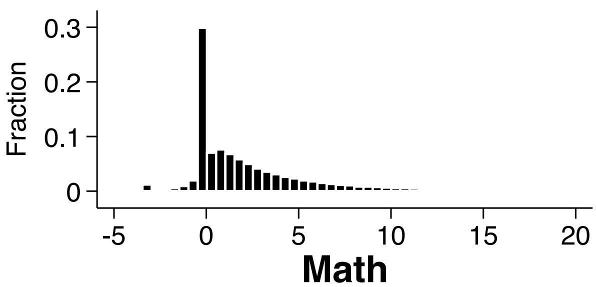
(c) Students with high academic performance



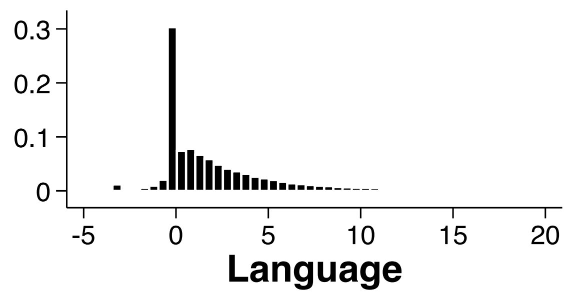
(d) Students with average academic performance

Notes: These figures present binscatter plots of "true test scores" (y-axis) and "predicted test scores" (x-axis) for different types of students. "True test scores" are observed test scores and "predicted test scores" were calculated using predetermined observable variables as predictors, combined using the estimated model in section 4 of the paper (equation 3). In these prediction exercises, we use the universe of test-takers – 1,929,654 students in the period 2005-2013 – and we delete 10% of observations in each school-year. Then, we proceed to predict test scores of the observations we deleted using the remaining 90% of students. This method allows us to evaluate the quality of our prediction. Students with low (high) academic performance are those below (above) the 25*th* (75*th*) percentile of the GPA distribution within a school-year.

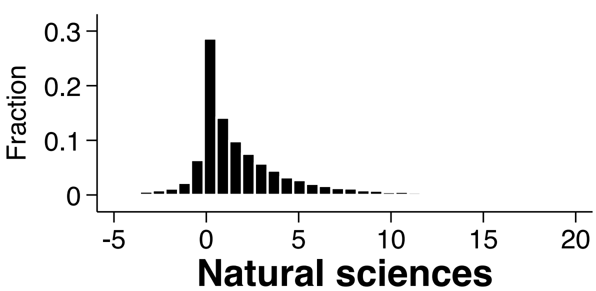
Figure A.9: Distribution of distortions by subject in 4*th* grade



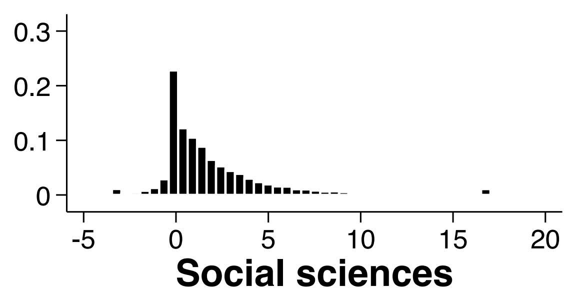
(a) 4*th* grade



(b) 4*th* grade



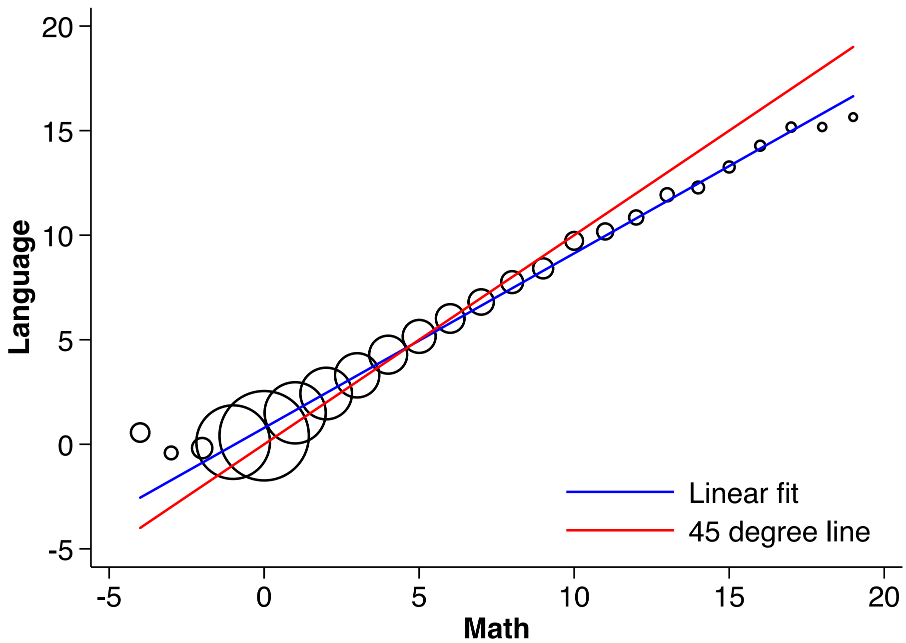
(c) 4*th* grade



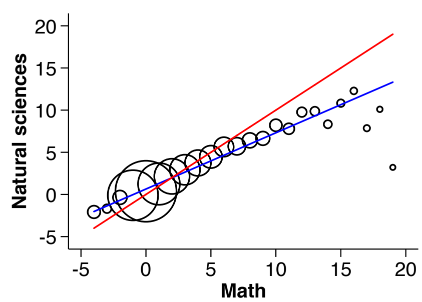
(d) 4*th* grade

Notes: We estimate distortions by subject of SIMCE using the methodology described in section 4 of the paper. Distortions in quality signals correspond to the average distortion in mathematics and language. We provide descriptive statistics for distortions by subject in Table A.1.

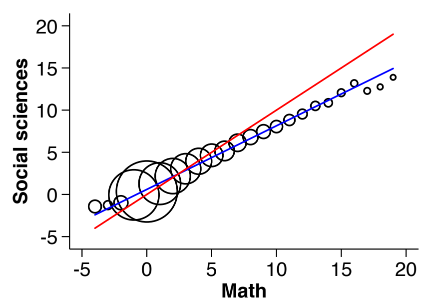
Figure A.10: Correlation between distortions in different tests



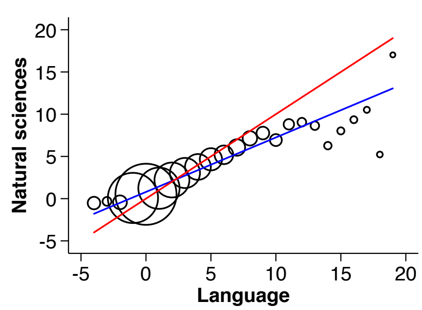
(a) Main tests



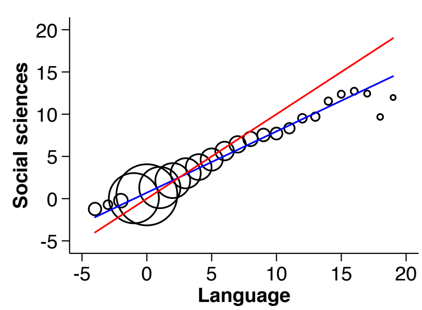
(b)



(c)



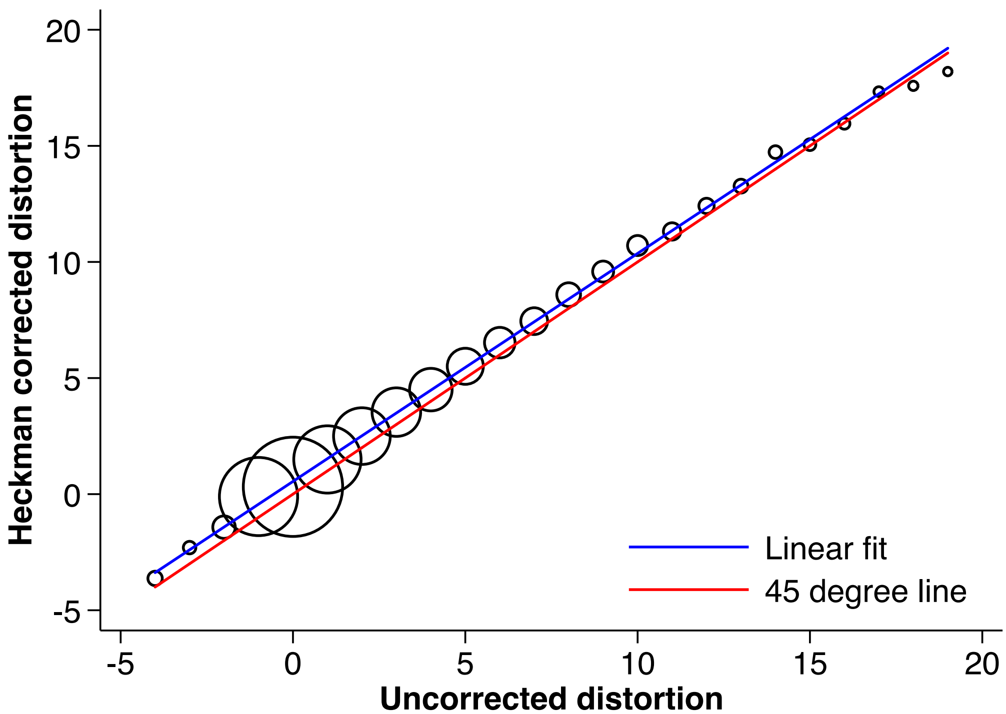
(d)



(e)

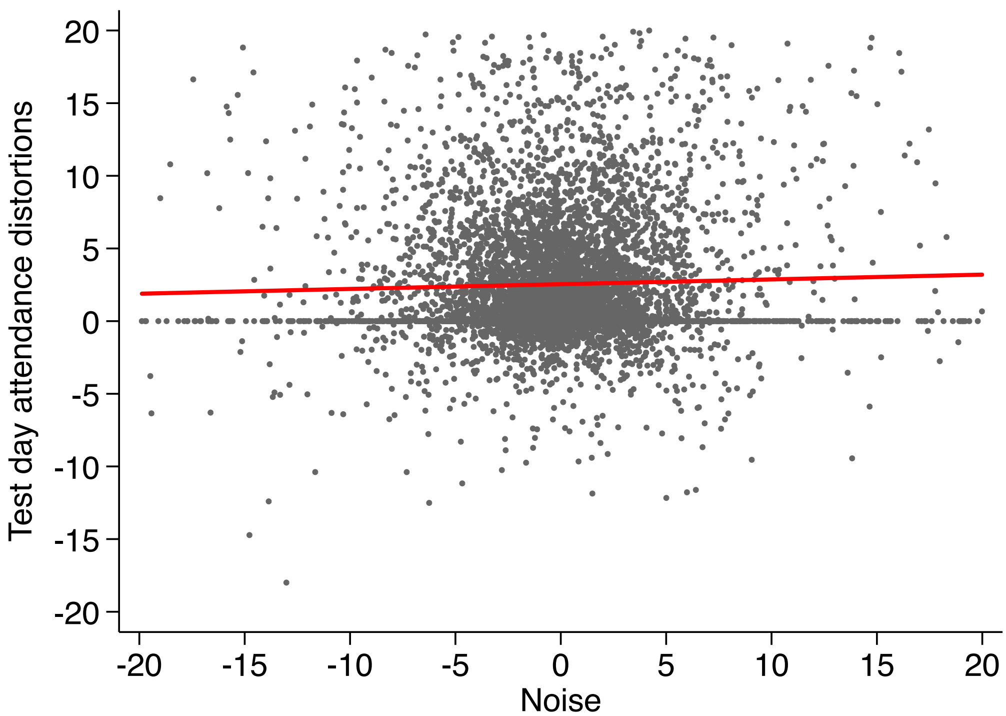
Notes: These figures displays the relationships between estimated distortions in test scores for different subjects of SIMCE.

Figure A.11: Heckman corrected distortions



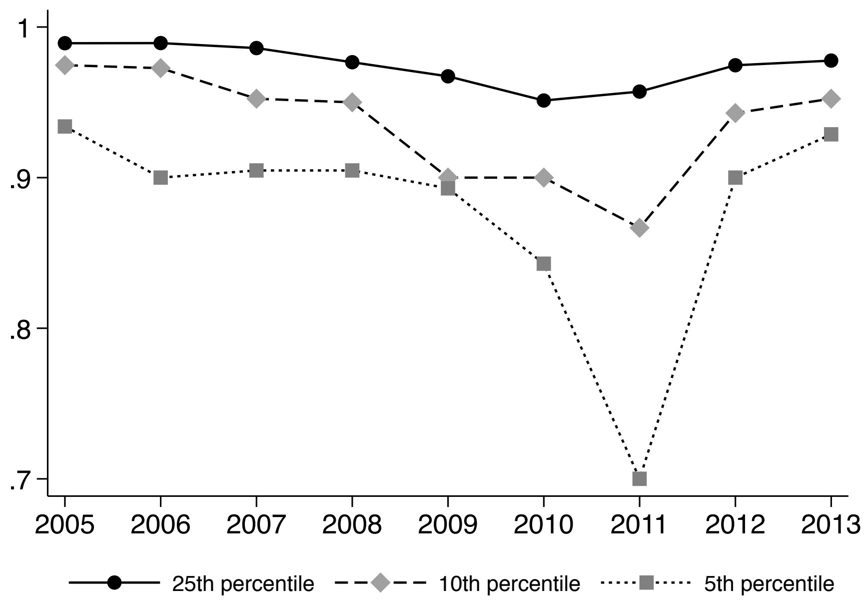
Notes: The excluded variable when calculating the Heckman corrected distortions is an indicator variable that takes the value of one for students that live outside of the municipality of the school.

Figure A.12: Distortions and noise

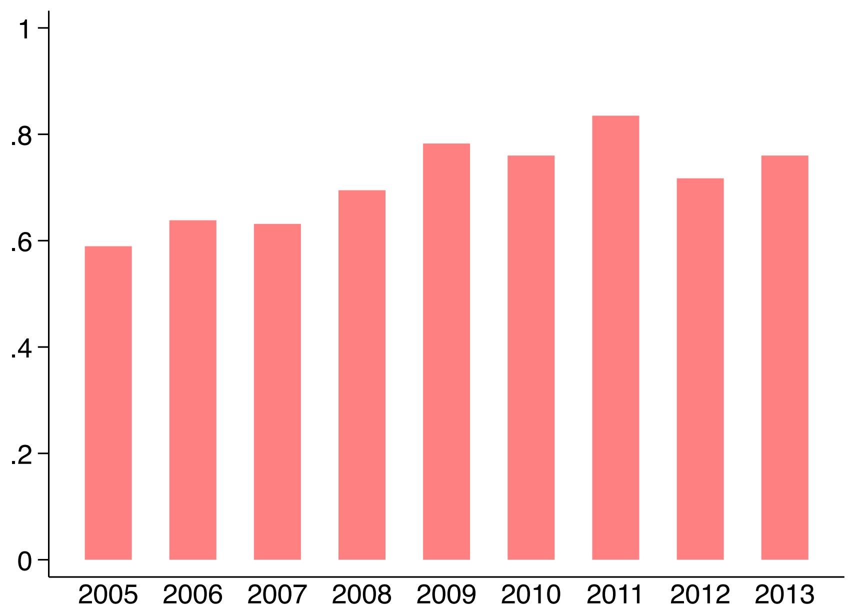


Notes: We construct a noise distribution for each school in our data using administrative estimates of noise in student-level test scores. These estimates are called "individual-level variability in test scores" and can be aggregated to construct measures of school-level noise following the method in Quality Education Agency (2013). This figure corresponds to a scatter plot showing the correlation between noise and distortions across schools as a linear fit. Each dot represents a school. The low correlation of 0.02 highlights that noise is mean independent of distortions. We conclude from this exercise that test day attendance represents a different (behavioral, non-statistical) margin that distorts quality signals.

Figure A.13: Distribution of rank correlations over time



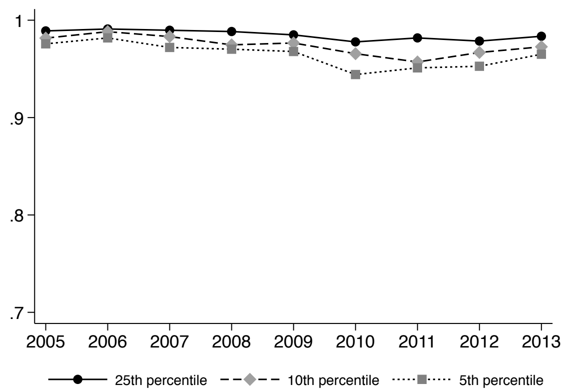
(a) Percentiles in rank correlation distribution $f(\rho_{mt})$



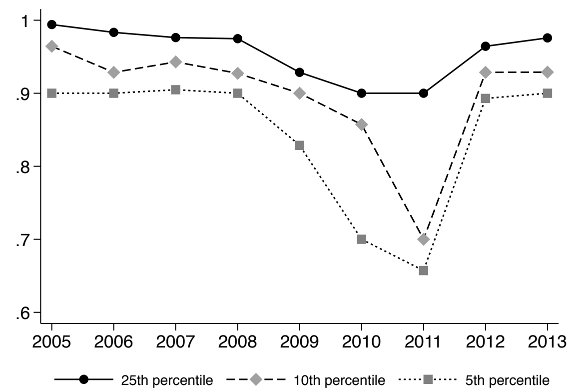
(b) Percentage of markets with changes in ranking (i.e., $\rho_{mt} < 1$)

Notes: Let ρ_{mt} be the rank correlation of distorted and undistorted quality in market m and year t. We observe approximately 210 markets every year.

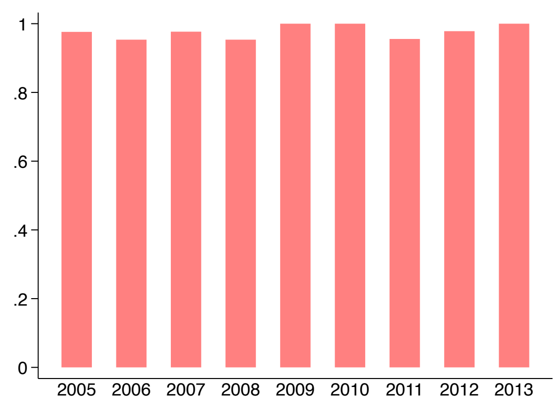
Figure A.14: Distribution of rank correlations by market type



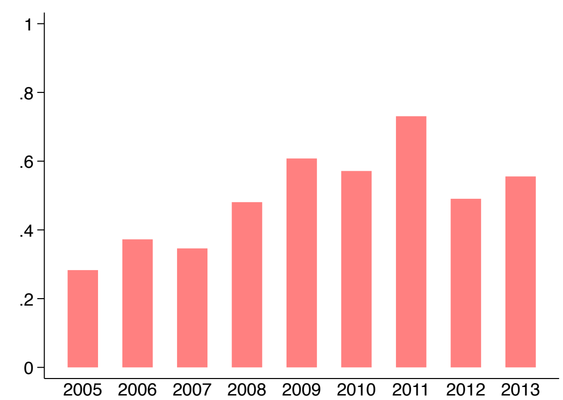
(a) Percentiles in large markets



(b) Percentiles in small markets



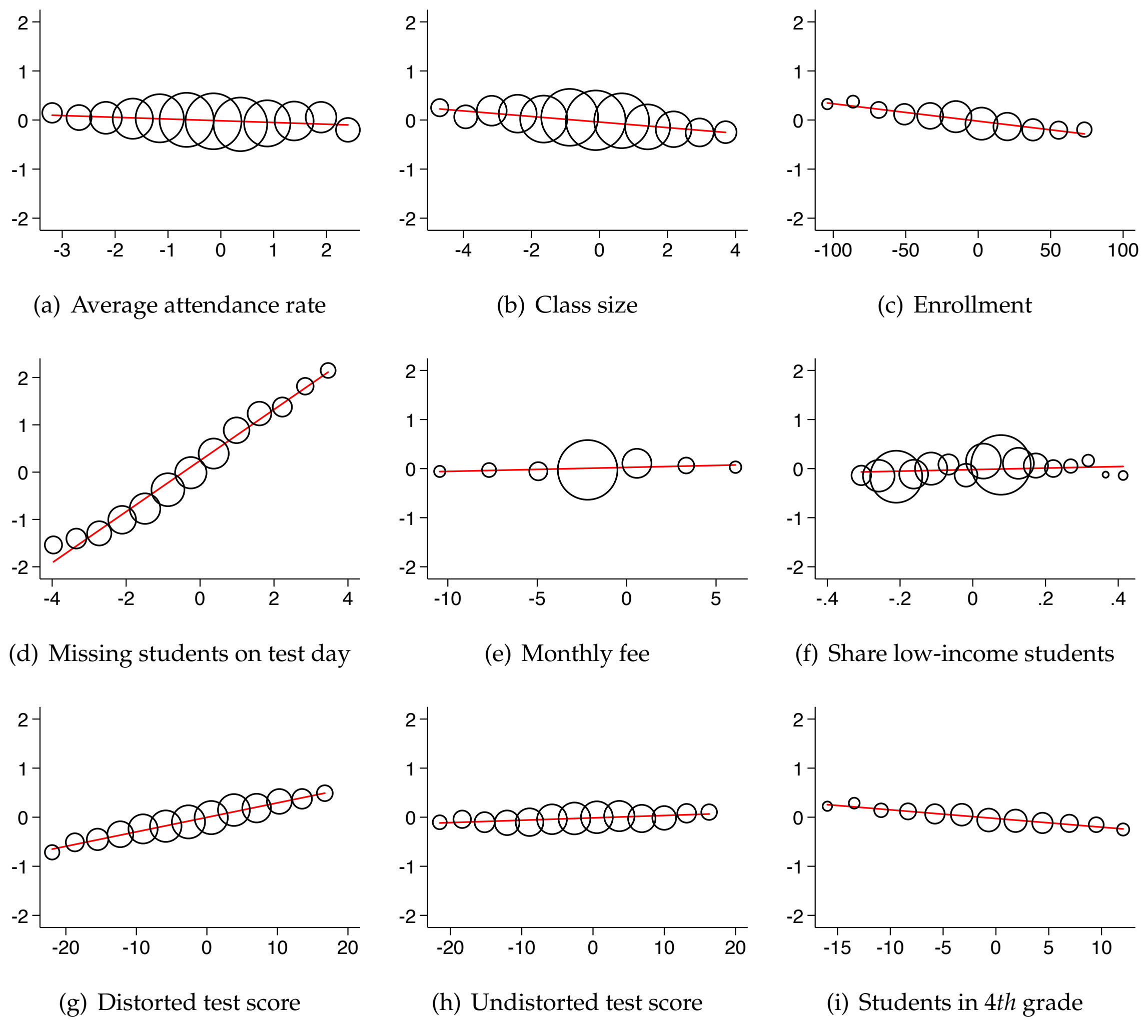
(c) Changes in ranking in large markets



(d) Changes in ranking in small markets

Notes: Let ρ_{mt} be the rank correlation of distorted and undistorted quality in market m and year t. We observe approximately 210 markets every year. "Percentiles in small/large markets" plot the percentiles in the rank correlation distribution $f(\rho_{mt})$ in market m and year t. "Changes in ranking in small/large markets" plot the percentage of markets with changes in ranking, i.e., $\rho_{mt} < 1$. Large (small) markets are defined as market-year observations with more (less) than 10 schools, the median number of schools.

Figure A.15: Distortions and school attributes



(a) Average attendance rate

(b) Class size

(c) Enrollment

(d) Missing students on test day

(e) Monthly fee

(f) Share low-income students

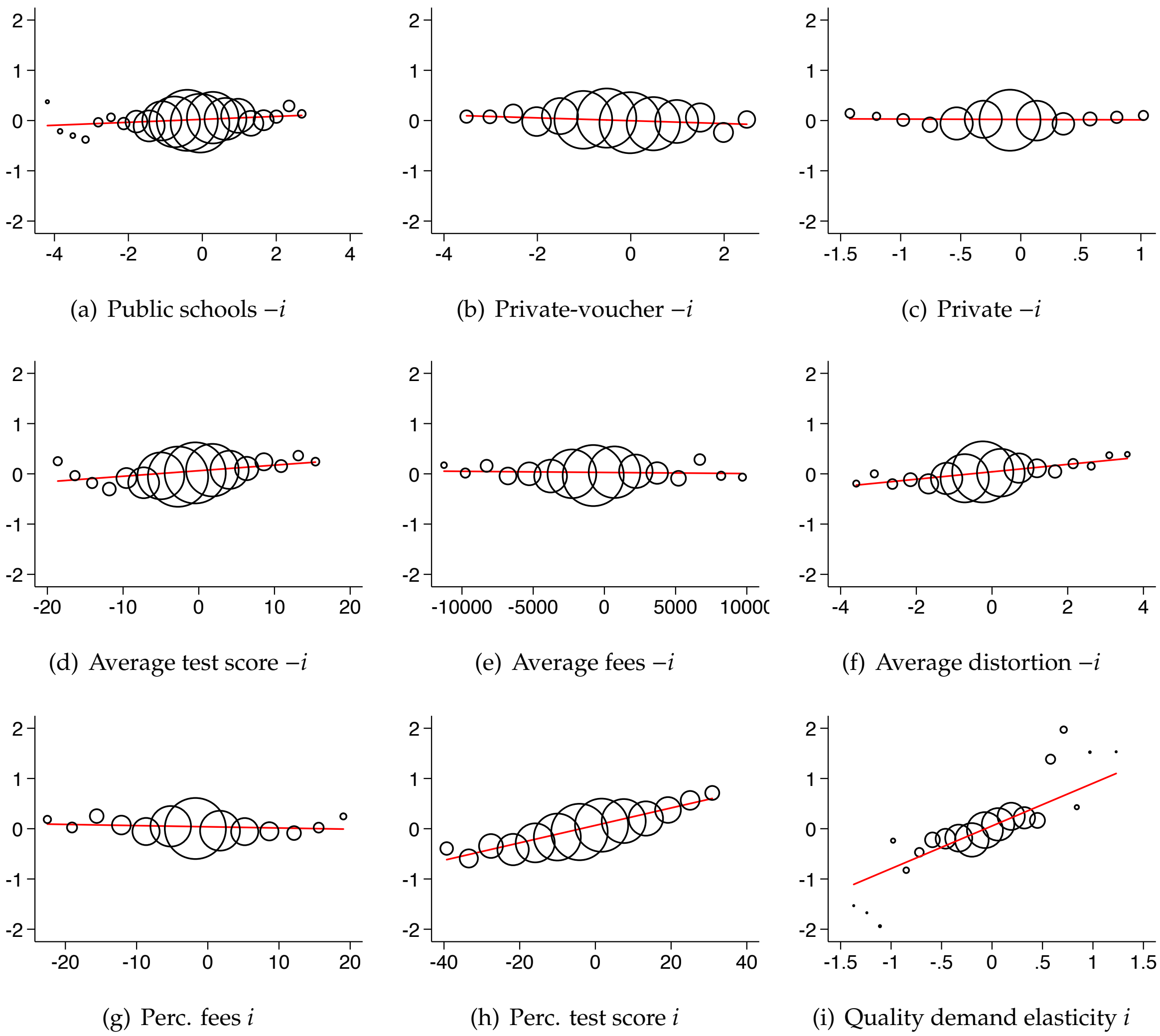
(g) Distorted test score

(h) Undistorted test score

(i) Students in 4*th* grade

Notes: These figures display the relationship between relevant school characteristics and distortions in quality signals. All variables have been residualized with school and year fixed effects. The size of markers indicates the number of students in it. The mean of distortion (y-axis) is 2.7 test score points.

Figure A.16: Distortions and attributes of schools within 3km



(a) Public schools $-i$ (b) Private-voucher $-i$ (c) Private $-i$

(d) Average test score $-i$ (e) Average fees $-i$ (f) Average distortion $-i$

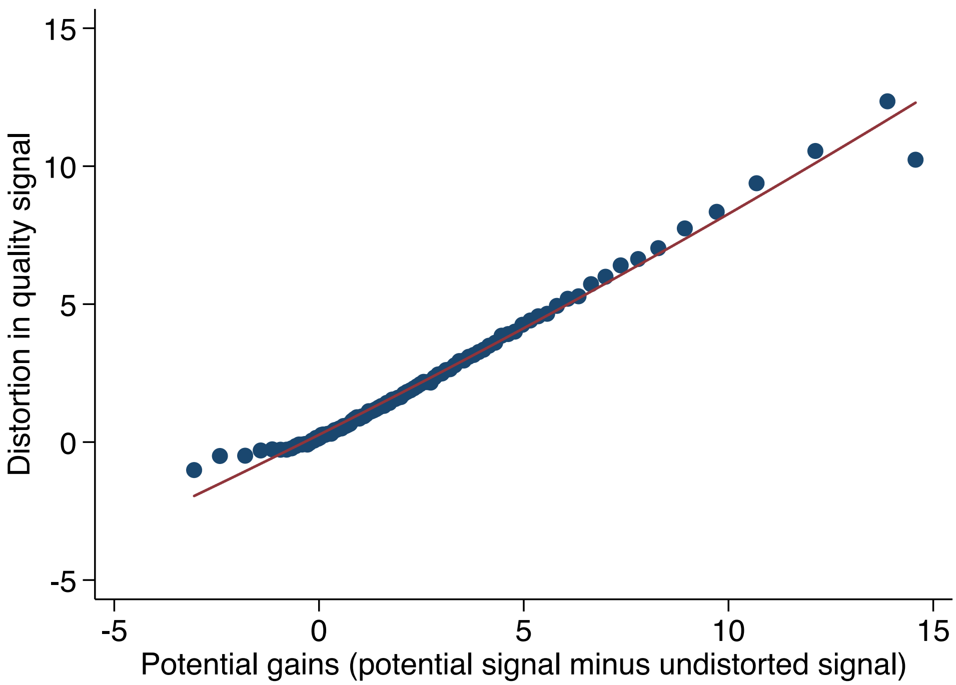
(g) Perc. fees i (h) Perc. test score i (i) Quality demand elasticity i

Notes: These figures display the relationship between relevant market characteristics and distortions in quality signals. All variables have been residualized with school and year fixed effects. The size of markers indicates the number of students in it. The mean of distortion (y-axis) is 2.7 test score points. Variables in panels (a) through (f) correspond to market aggregates excluding the reference school. Quality demand elasticities in panel (i) are calculated using the sample and estimates from the school choice model in section 5, as:

$$\eta^q_{jmt} = \frac{\partial s_{jmt}}{\partial q_{jmt}} \frac{q_{jmt}}{s_{jmt}} = \left(\sum_r \pi^r_{mt} \frac{1}{N^r_{mt}} \sum_{i \in \mathcal{I}^r_{mt}} \frac{\partial P^r_{ijmt}(d^r, \hat{\delta}^r, \hat{\beta}^r_d)}{\partial q_{jmt}} \right) \frac{q_{jmt}}{s_{jmt}}$$

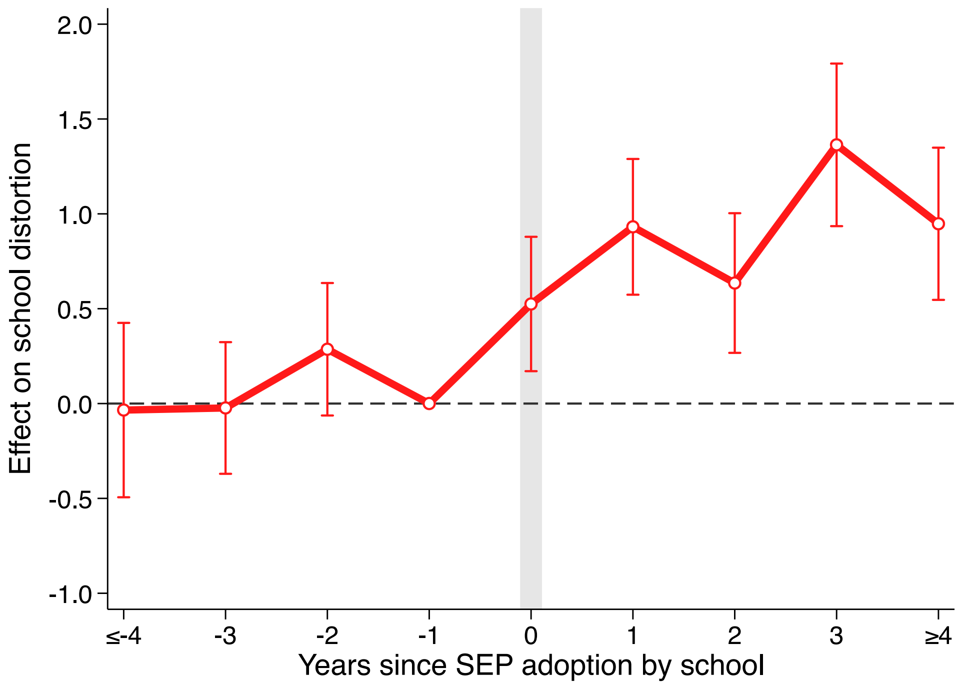
where π^r_{mt} is the share of households of type r in market m and year t, while N^r_{mt} and $\mathcal{I}^r_{mt}$ are the number and the set of such households respectively. The expression in brackets is thus simply a type-share-weighted average of the partial derivative of choice probabilities for school j with respect to quality. In the plot, both variables are residualized by removing school and year fixed effects.

Figure A.17: Potential gains and distortions



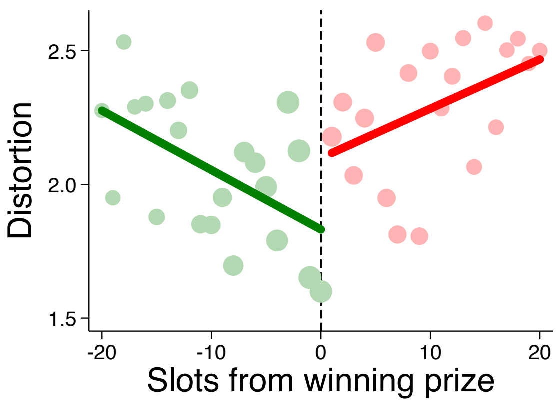
Notes: This figure displays a binned scatter plots of distortions on potential gains from having students not attend on test days. We compute potential gains as the difference between the average predicted school test score if the bottom 10 percent of the student GPA performance distribution does not take the test and the average predicted school test score if all students in the class take the test. The latter is what we call schools' undistorted quality signals. Both variables have been residualized with school and year fixed effects.

Figure A.18: Change in distortions around adoption of SEP program

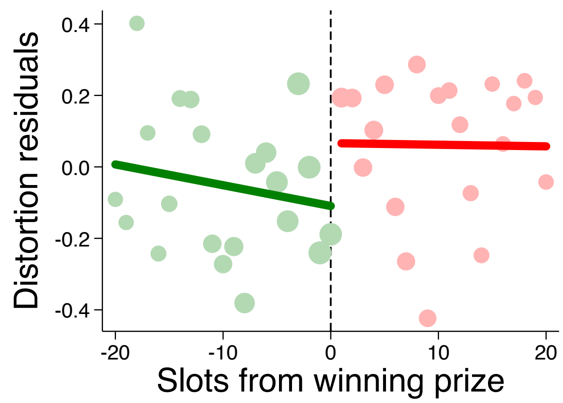


Notes: Event study analysis in equation (1) to test for the effect of affiliation to the SEP vouchers program on distortions in quality signals (y-axis). The x-axis displays years since a school adopts to the SEP program. Dots indicate coefficients for the effect of each year around the event on distortions in quality signals. The coefficient on the year before SEP adoption is normalized to zero. Clustered standard errors at the school level are displayed in brackets. We find that distortions in quality signals increase after SEP adoption.

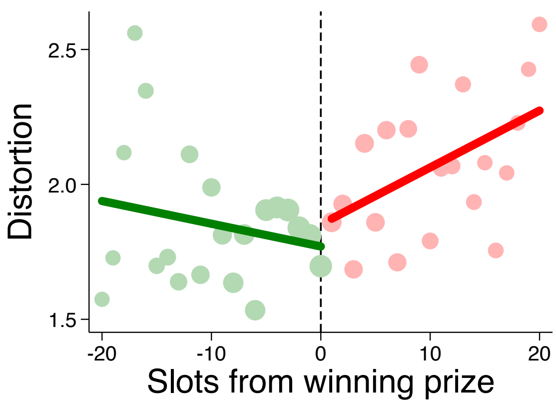
Figure A.19: Monetary incentives for teachers



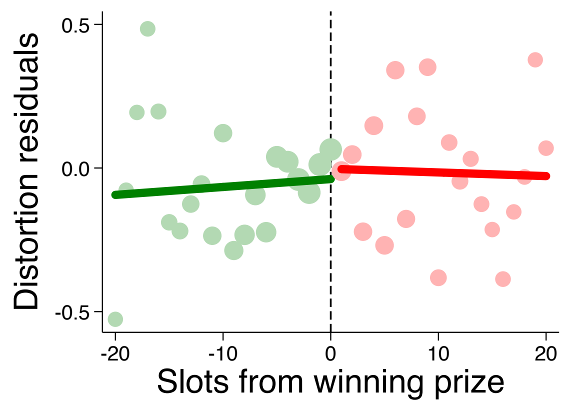
(a) Distortion in year $t + 1$



(b) Distortion in year $t + 1$



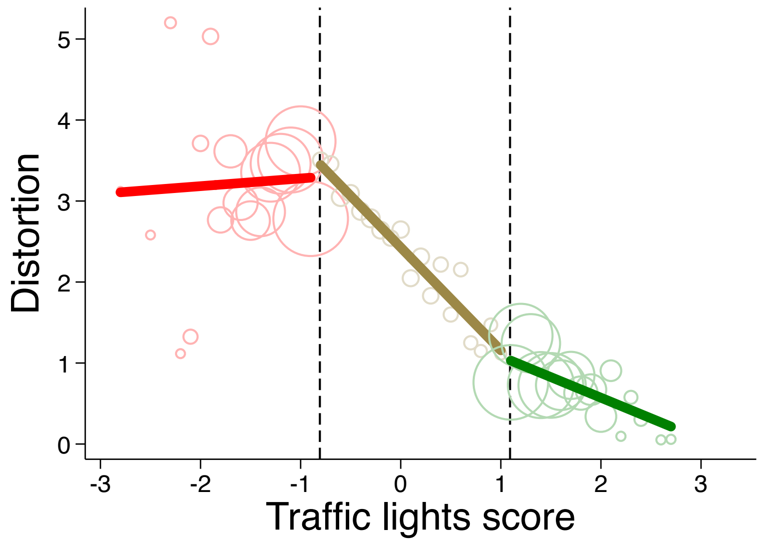
(c) Distortion in year $t + 2$



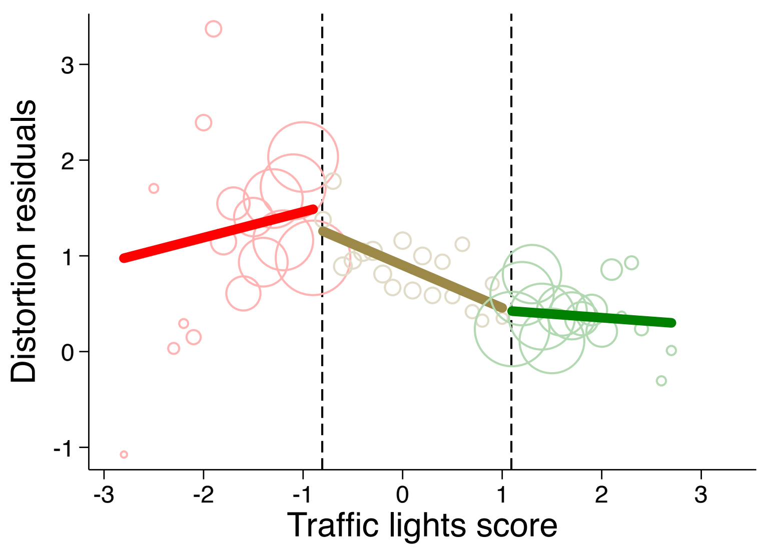
(d) Distortion in year $t + 2$

Notes: Regression kink design in equation (2) to test for the effect of monetary teacher incentives on distortions in quality signals (y-axis). The x-axis represents a measure of the probability of winning the prize (i.e. teacher bonuses). Schools to the left (right) of the thresholds won (did not win) the prize in the previous tournament. We present more details about this public program in section 2 of the paper. Left panels correspond to changes in the slope without controls while right panels control for a set of school fixed effects. The null hypothesis of incentives affecting distortions implies an "inverted V" relationship between "slots from winning prize" and distortions around the kink. We reject the hypothesis that teacher incentives cause distortions in quality signals.

Figure A.20: "Educational Traffic lights" policy



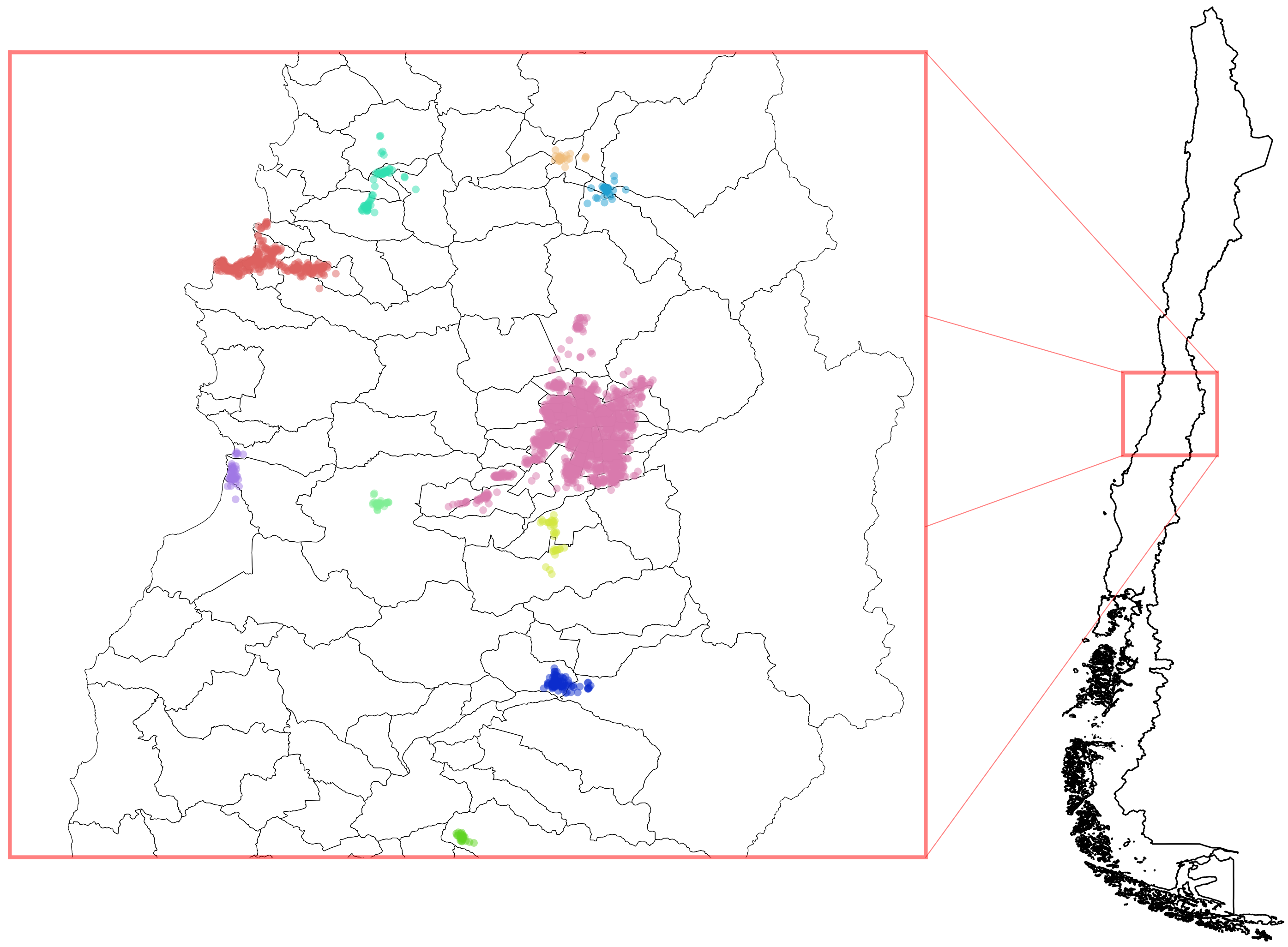
(a) Distortion in year 2010



(b) Distortion in year 2010

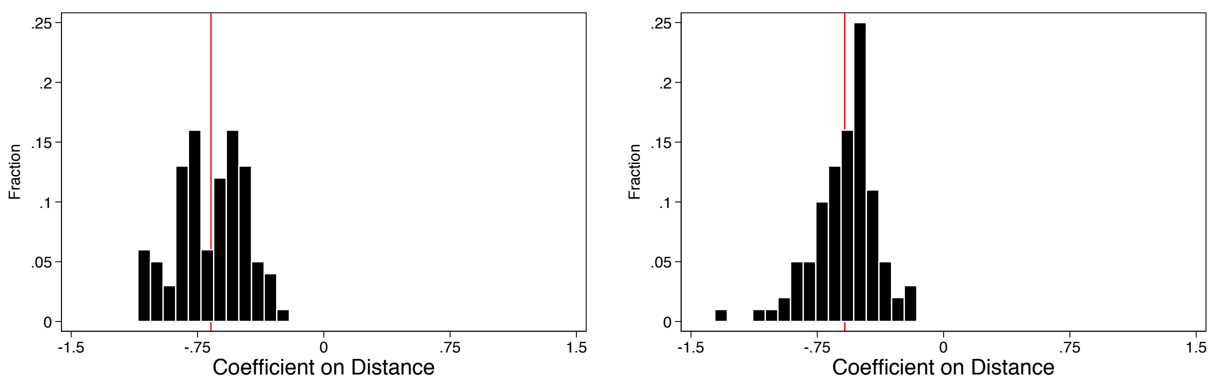
Notes: Regression kink design in equation (3) to test for the hypothesis of manipulation of test scores to be classified in a "higher" category. The x-axis represents school scores which fully determines their category. We present more details about the policy in section 2 of the paper. The null hypothesis of manipulation implies an "inverted V" relationship between school scores and distortions in quality signals. The upper panel corresponds to the test without controls while the lower panel controls for a basic set of pre-determined school characteristics. We strongly reject the hypothesis of manipulation of test scores for the school to be classified in a higher category.

Figure A.21: Market definition



Notes: See description of Table A.2 for details about market definition. This map plots the ten largest markets in the most populated area of the country.

Figure A.22: Estimated coefficients on distance from the first stage

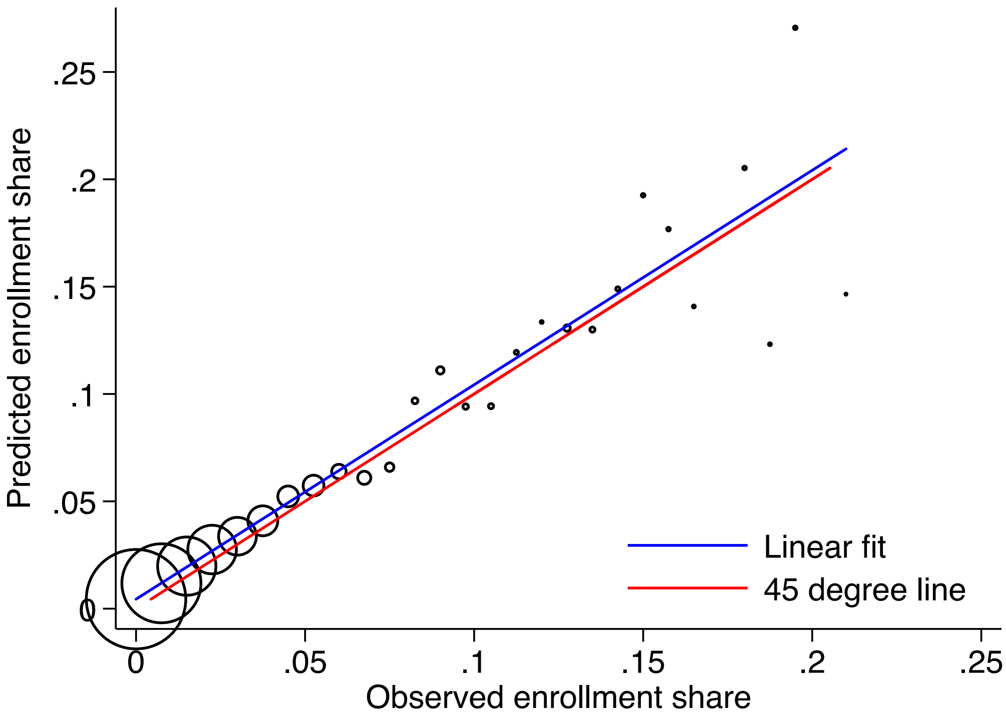


(a) Low-income students, distance

(b) High-income students, distance

Notes: These figures display resulting estimates for β_d^r from the first stage of the school choice model. Each observation is the estimated coefficient for an estimating cell comprised by a market, year and household type. The red line indicates the average coefficient.

Figure A.23: Observed and predicted school enrollment

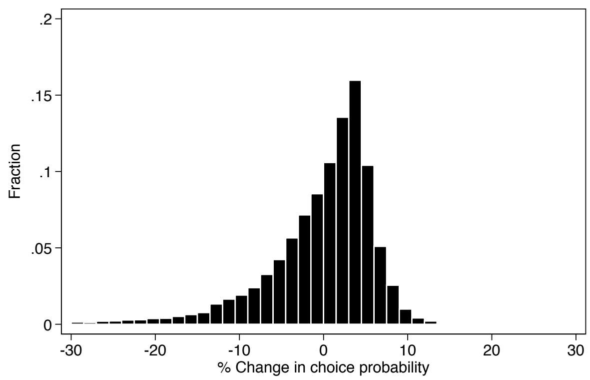


Notes: This figure displays the relationship between observed school enrollment shares and predicted school enrollment shares using model estimates. Predicted enrollment shares are calculated as:

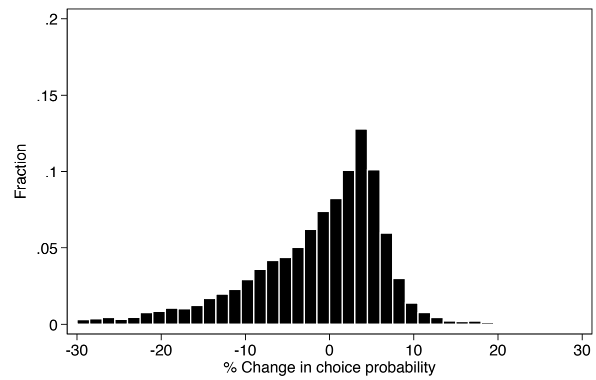
$$s_{jmt}(\hat{\delta}, \hat{\beta}_d) = \sum_r \pi^r_{mt} \frac{1}{N^r_{mt}} \sum_{i \in \mathcal{I}^r_{mt}} P^r_{ijmt}(d^r, \hat{\delta}^r, \hat{\beta}^r_d)$$

where π^r_{mt} is the share of households of type r in market m and year t, while N^r_{mt} and $\mathcal{I}^r_{mt}$ are the number and the set of such households respectively. The expression is thus simply a type-share-weighted average of average choice probabilities for school j.

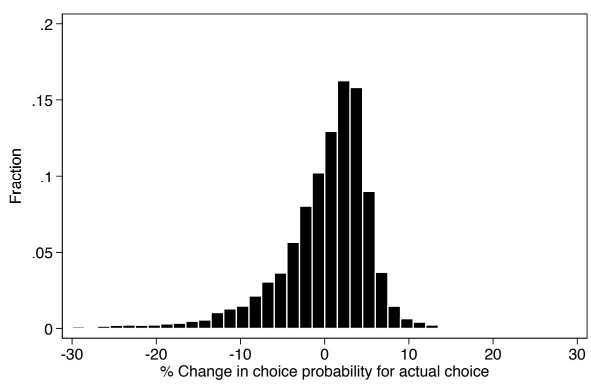
Figure A.24: Changes in choice probabilities



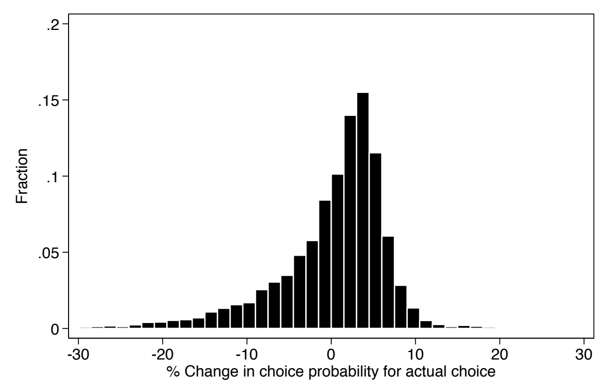
(a) Low-income students, all schools



(b) High-income students, all schools



(c) Low-income students, chosen school



(d) High-income students, chosen school

Notes: These figures display change in school choice probabilities between the counterfactual and baseline scenarios we analyze. Each observation is the percentage change in the choice probability of a school by a household in the estimating dataset. Panels (a) and (b) include results for all schools in the dataset, while panels (c) and (d) focus only on schools chosen by household in the baseline scenario.

Table A.1: Descriptive statistics for distortions by subject

	Obs.	Mean	St. Dev.	Min	Max	Years
Mathematics	60,741	2.7	4.4	-3.5	23.9	2005–2013
Language	60,760	2.6	4.4	-3.4	23.8	2005–2013
Natural sciences	5,902	2.1	3.9	-7.8	20.6	2008, 2010
Social sciences	10,033	2.1	3.2	-3.5	17.0	2009

Notes: Distortions are measured in test score points and we estimated them using the methodology described in section 4. See Figure A.2 for a timeline of standardized tests.

Table A.2: School markets as connected components

	3km	4km	5km	6km	7km	8km	9km	10km
Markets	451	413	380	348	322	295	273	251
Markets with more than 1 schools	262	248	233	219	208	196	191	181
Markets with more than 5 schools	106	104	99	93	90	88	86	86
Markets with more than 10 schools	63	63	60	55	52	49	48	50
Markets with more than 20 schools	36	36	33	31	30	29	28	29

Notes: Let **A** be a $N \times N$ matrix representing the network of N =5,416 urban schools in Chile in the period 2005–2013. In network theory, **A** is referred to as adjacency matrix. This adjacency matrix represents an undirected network, i.e., **A** is a symmetric matrix. The element $A(i, j)$ in this adjacency matrix takes the value of one if school i and j are closer than κ kilometers from each other and zero otherwise. A "component" or "connected component" of **A** is a sub-network in which any two schools are connected to each other through some other school, i.e., we can always find a "path" that connects any two pair of schools in the sub-network. A market is defined as a connected component of **A**. In the paper, we use κ = 5 (highlighted in gray), but results are robust to different definitions.

Table A.3: Summary statistics for estimation of school choice model

Variable		Mean	St. dev.	p10	p50	p90
Students	In sample	1,009	844	324	665	2,446
	Coverage rate	0.33	0.12	0.17	0.31	0.48
Schools	In sample	63	62	19	45	134
	Coverage rate	0.92	0.13	0.72	0.97	1.00
Low-income students	In sample	479	391	166	323	1,184
	Sample share	0.49	0.11	0.35	0.50	0.60

Notes: This table displays market-level summary statistics for the sample we use to estimate the school choice model. This sample includes 25 markets in the period 2011–2014. For the number of students and schools per market, we provide summary statistics in levels and coverage rate of the complete market. For the number of low-income students, we provide summary statistics of levels and their share over the sample market size.

Table A.4: IV results from the second stage of school choice model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All			Low-income students			High-income students		
Fee	-0.003***	-0.004***	-0.006***	-0.006***	-0.007***	-0.010***	-0.002***	-0.003***	-0.005***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)
Quality	0.012***	0.019***	0.019***	0.004**	0.011***	0.015***	0.021***	0.028***	0.023***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Religious			-0.054**			-0.086***			-0.019
			(0.024)			(0.030)			(0.029)
Gender constraint			0.148***			0.121**			0.161***
			(0.047)			(0.059)			(0.055)
Public			0.089***			0.229***			-0.061
			(0.031)			(0.040)			(0.039)
SEP school			-0.325***			-0.533***			-0.587***
			(0.065)			(0.094)			(0.066)
Market-year F.E.	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	10,774	10,774	10,774	5,335	5,335	5,335	5,439	5,439	5,439
First stage tests									
F-test fee	1566.15	2031.73	395.15	484.55	582.07	73.91	1285.81	1593.35	329.15
F-test quality	70.62	17.69	15.03	33.81	9.68	8.17	36.50	8.03	6.86
Cragg-Donald EV	283.97	232.38	203.91	146.05	127.12	101.57	139.36	106.47	98.52

Notes: Instrumental variable estimates. We use two sets of instruments: (i) the amount awarded by school vouchers, mean fixed characteristics of rivals in the market (i.e. BLP instruments) and rivals market wages are used as instruments for schools fees; and (ii) a linear and quadratic term on county-specific temperature and the residual of a regression of being awarded a SNED prize in the previous period on lagged school quality are use as instruments for school quality. F-tests are computed separately for each first stage for the respectively excluded instruments. All regressions are weighted by school enrollment. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.5: IV results from the second stage of school choice model - First stage for school fees

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		All			Low-income students			High-income students	
Rivals' characteristics									
Religious	-19.876**	79.768*	100.580	-11.500	68.481	11.209	-26.668**	85.263	195.296*
	(7.810)	(46.975)	(70.650)	(10.365)	(57.551)	(79.821)	(11.316)	(71.460)	(108.364)
Gender constraint	-80.777***	463.861***	780.010***	-90.315***	513.487***	617.324**	-62.196**	418.205**	922.088***
	(19.195)	(121.244)	(222.163)	(24.978)	(149.890)	(257.538)	(28.246)	(182.350)	(335.240)
Public	-17.294***	233.888***	-83.801	-9.406	-114.488**	-75.163	-21.657**	614.363***	-68.761
	(6.075)	(45.861)	(61.668)	(7.921)	(48.773)	(70.400)	(9.017)	(74.814)	(94.836)
Teacher wages									
Average hourly wage	0.094***	0.123***	0.135***	0.060	0.078*	0.089**	0.131***	0.170***	0.184***
	(0.031)	(0.032)	(0.030)	(0.042)	(0.043)	(0.041)	(0.045)	(0.047)	(0.045)
Voucher									
Baseline	-1.335***	-1.473***	-1.454***	-0.975***	-1.177***	-1.166***	-1.535***	-1.627***	-1.603***
	(0.031)	(0.030)	(0.029)	(0.059)	(0.055)	(0.054)	(0.037)	(0.036)	(0.035)
SEP school	0.062*	-0.080**	0.007	-0.093***	-0.217***	-0.077**	0.210***	0.056	0.090*
	(0.032)	(0.033)	(0.032)	(0.034)	(0.035)	(0.033)	(0.052)	(0.052)	(0.049)
SEP share	-10.034***	-9.616***	-1.556***	-9.927***	-9.825***	-0.856	-10.113***	-9.368***	-2.342***
	(0.185)	(0.187)	(0.546)	(0.235)	(0.241)	(0.636)	(0.279)	(0.279)	(0.815)
Temperature									
Linear	4.167***	2.684	0.621	3.980**	3.871	0.808	4.284**	0.709	-0.311
	(1.308)	(4.491)	(4.381)	(1.778)	(6.120)	(5.784)	(1.874)	(6.490)	(6.466)
Quadratic	-0.261***	-0.163	-0.095	-0.256***	-0.214	-0.110	-0.260***	-0.088	-0.058
	(0.054)	(0.179)	(0.174)	(0.074)	(0.243)	(0.230)	(0.078)	(0.259)	(0.258)
SNED program									
Prize residual	-5.773***	-4.227***	-5.363***	-3.588**	-2.150	-3.406**	-7.605***	-5.997***	-7.023***
	(1.211)	(1.159)	(1.143)	(1.560)	(1.499)	(1.472)	(1.816)	(1.727)	(1.698)
Controls	No	No	Yes	No	No	Yes	No	No	Yes
Market-Year F.E.	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	10,774	10,774	10,774	5,335	5,335	5,335	5,439	5,439	5,439
R-squared	0.685	0.729	0.739	0.665	0.716	0.730	0.706	0.748	0.758

Notes: All regressions are weighted by school enrollment. Columns 3, 6 and 9 include other school attributes in the corresponding second stage specifications, namely indicators for schools being religious, public, gender constrained or part of the SEP program. Results not reported in this table. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.6: IV results from the second stage of school choice model - First stage for school quality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		All			Low-income students			High-income students	
Rivals' characteristics									
Religious	1.173	12.207	62.650**	2.547	9.472	54.491	-0.046	14.997	71.121*
	(3.687)	(20.849)	(29.730)	(5.269)	(29.957)	(42.769)	(5.157)	(29.337)	(41.727)
Gender constraint	-42.162***	-26.845	375.535***	-45.462***	-27.096	372.044***	-38.185***	-26.725	377.581***
	(8.451)	(46.254)	(77.523)	(12.077)	(65.914)	(110.647)	(11.852)	(65.569)	(109.687)
Public	-7.425***	322.369***	-34.034	-6.927*	321.522***	-46.263	-7.653**	324.807***	-20.461
	(2.729)	(22.419)	(30.724)	(3.881)	(32.046)	(43.895)	(3.834)	(31.777)	(43.497)
Teacher wages									
Average hourly wage	0.060***	0.064***	0.072***	0.059***	0.062***	0.071***	0.061***	0.067***	0.074***
	(0.013)	(0.014)	(0.013)	(0.018)	(0.019)	(0.019)	(0.018)	(0.019)	(0.019)
Voucher									
Baseline	-0.212***	-0.247***	-0.234***	-0.183***	-0.231***	-0.219***	-0.227***	-0.254***	-0.241***
	(0.009)	(0.009)	(0.009)	(0.016)	(0.016)	(0.015)	(0.012)	(0.012)	(0.011)
SEP school	0.715***	0.604***	0.625***	0.720***	0.615***	0.636***	0.709***	0.593***	0.614***
	(0.018)	(0.017)	(0.017)	(0.025)	(0.024)	(0.024)	(0.025)	(0.025)	(0.024)
SEP share	-4.663***	-4.149***	-0.157	-4.694***	-4.200***	-0.177	-4.628***	-4.092***	-0.134
	(0.082)	(0.079)	(0.202)	(0.115)	(0.113)	(0.289)	(0.117)	(0.113)	(0.286)
Temperature									
Linear	4.151***	1.492	0.620	4.134***	1.727	0.620	4.166***	1.196	0.554
	(0.494)	(1.630)	(1.489)	(0.712)	(2.339)	(2.115)	(0.686)	(2.298)	(2.122)
Quadratic	-0.208***	-0.091	-0.064	-0.206***	-0.105	-0.067	-0.209***	-0.076	-0.059
	(0.021)	(0.066)	(0.060)	(0.030)	(0.094)	(0.085)	(0.029)	(0.093)	(0.085)
SNED program									
Prize residual	1.713***	3.029***	2.412***	1.861***	3.143***	2.492***	1.598**	2.935***	2.349***
	(0.486)	(0.447)	(0.431)	(0.687)	(0.636)	(0.614)	(0.689)	(0.634)	(0.612)
Controls	No	No	Yes	No	No	Yes	No	No	Yes
Market-Year F.E.	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	10,774	10,774	10,774	5,335	5,335	5,335	5,439	5,439	5,439
R-squared	0.391	0.474	0.512	0.367	0.452	0.492	0.412	0.493	0.529

Notes: All regressions are weighted by school enrollment. Columns 3, 6 and 9 include other school attributes in the corresponding second stage specifications, namely indicators for schools being religious, public, gender constrained or part of the SEP program. Results not reported in this table. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.7: OLS results from the second stage of school choice model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All sample			Low-income students			High-income students		
Fee	-0.004***	-0.004***	-0.003***	-0.009***	-0.010***	-0.009***	-0.002***	-0.001***	-0.003***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)
Quality	0.010***	0.011***	0.011***	0.003***	0.005***	0.006***	0.017***	0.017***	0.016***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Religious			-0.040*			-0.074**			0.000
			(0.023)			(0.029)			(0.028)
Gender constraint			0.233***			0.237***			0.222***
			(0.043)			(0.053)			(0.050)
Public			0.039			0.126***			-0.049
			(0.026)			(0.032)			(0.033)
SEP school			0.012			0.028			-0.247***
			(0.042)			(0.054)			(0.045)
Market-Year F.E.	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	11,041	11,041	11,041	5,461	5,461	5,461	5,580	5,580	5,580
R-squared	0.025	0.370	0.373	0.072	0.488	0.494	0.080	0.561	0.567

Notes: All regressions are weighted by school enrollment. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.8: Schools in corrupt municipalities have larger distortions

Dependent variable is distortions (in test score points)

	Years with transfers	Before audits revealed	After audits revealed
	(1)	(2)	(3)
Irregular payments	0.04***	0.05***	0.02
	(0.01)	(0.02)	(0.02)
Government transfers	0.08***	0.09***	0.07***
	(0.01)	(0.02)	(0.02)
Schools	2,345	2,283	2,239
Municipalities	76	76	76
Observations	11,834	7,588	4,246

Notes: All variables have been normalized. All regressions are weighted by the inverse of the size of the confidence interval of distortions to account for estimation of the dependent variable. Audits in 76 randomly chosen municipalities were implemented by the Comptroller General of Chile to disclose irregular payments from government transfers. The time of disclosure of irregular payments was May of 2012. "Years with transfers" correspond to the period 2008–2013. Column 2 restricts attention to years 2008–2012, and column 3 restricts attention to years 2012–2013. Robust standard errors in parentheses. Significance level: *** p<0.01, ** p<0.05, * p<0.1.

Table A.9: Differences-in-differences of audits

Dependent variable is distortions (in test score points)

	All schools		Schools in audited municipalities	
	(1)	(2)	(3)	(4)
Audit × Post	0.07	0.04		
	(0.05)	(0.03)		
Corrupt × Post			-0.17**	-0.12**
			(0.07)	(0.05)
Post	-0.04*	-0.04**	0.09	0.06*
	(0.02)	(0.02)	(0.06)	(0.03)
Mean of dep. variable	2.9	2.9	3.0	3.0
School-level controls	No	Yes	No	Yes
Municipality F.E.	Yes	Yes	Yes	Yes
Municipalities	344	344	76	76
Schools	7,357	7,357	2,239	2,239
Observations	40,705	37,448	12,865	11,834

Notes: These regressions restrict attention to the period in which the government transferred monetary resources to be spent under the *Subvención Escolar Preferencial* program (2008–2013). All regressions are weighted by the inverse of the size of the confidence interval of distortions to account for estimation of the dependent variable. Audits in 76 randomly chosen municipalities were implemented by the Comptroller General of Chile to disclose "irregular" expenditures of government transfers. The time of disclosure of irregular payments was May of 2012. The post period are years 2012 and 2013. The "Corrupt" indicator takes the value of one if a municipality has more than 10 percent of the government transfers under "irregular payments." More about irregular payments can be found in CIPER (2012). Standard errors clustered at the municipality level in parentheses. Significance level: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.