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 school quality is often gathered through standardized testing. However, the use of these tests has been controversial because of behavioral responses that could distort performance measures. 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 indicate that undesirable responses to test-based accountability systems may impose significant costs on educational markets.

Keywords: accountability, schools, quality, disclosure, choice

JEL Codes: I20, L15

†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 consumer choices. In education, information on school quality is often measured via standardized tests. The use of these tests to assess school performance 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 high-stakes testing might generate undesirable behavioral responses that introduce distortions in the performance metric itself and thus the accomplishment of its goals (Neal, 2013). This argument stems from Holmstrom and Milgrom (1991), who underscore the role of hidden actions in producing changes in the observed outcome (i.e. test scores) without necessarily improving the real outcome of interest (i.e. learning). Despite increasing evidence of undesirable behavioral responses, quantification of these potential distortions and their consequences is lacking.

How large are these distortions in school quality signals? What are the market consequences of these distortions? We study one of the most developed accountability systems in the world – Chile’s market-oriented educational system (Figlio and Loeb, 2011)– and show that behavioral responses are in place and distort key performance metrics. Distortions are large and have significant consequences on school choice and the allocation of public programs. The government relies on standardized testing to generate school-specific quality metrics. These measures are used not only for quality assessment and performance evaluation, but also as a disclosure system in school choice. These features make Chile an ideal setting to quantify the consequences of behavioral responses to accountability systems.

The analysis proceeds in four steps. First, we show that low-performing students are more likely to be absent on test days relative to other students. Using national administrative data on Chilean school children, we compare daily attendance of test takers (fourth graders) and non-takers (third graders) within schools on test and non-test days. High performing test-takers increase their attendance on test days by 0.18 standard deviations more than low performing test-takers. This result suggests that a behavioral response to standardized testing is at work. However, the degree of student non-representativeness varies considerably across schools.

Second, we use a multiple imputation method to predict the test scores of absent students and thus the associated distortions in school quality signals.\(^1\) We find average distortions in the system to be sizable: 0.1 standard deviations of school test scores. Distortions vary widely across schools, but are persistent within schools over time. To better understand these distortions, we construct a panel dataset for all Chilean schools during the period 2005–2013. Public, low quality, and for-profit schools display larger average distortions. We find some evidence of larger distortions in more competitive markets. In particular, schools facing more

\(^1\)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).
quality-elastic consumers display larger distortions in quality signals. In contrast, we find no evidence for potential perverse incentives by teacher performance programs in place.

Third, we estimate a school choice model to quantify the implications of these distortions. We find that providing undistorted school quality information would likely induce three percent of students to switch schools. To estimate the model, we use geocoded addresses of 100,000 students and 1,500 schools, and estimate a discrete choice model in which households trade-off school quality and distance. For identification, we exploit quasi-experimental variation in government programs, climate-induced variation in test scores, 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: attendance on test days. This type of behavioral response has not been found in the U.S., where the most common response has been the selective assignment of students to special education programs (Jacob, 2005; Rockoff and Turner, 2010; Figlio and Loeb, 2011). Second, we propose and implement 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 effect of distortions on school choice and the allocation of public programs and thus quantify the market consequences of these behavioral responses. While we implement our analysis in the Chilean educational market, 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 create incentives for undesirable behavioral responses (Dranove and Jin, 2010). Moreover, whenever quality signals generated by the disclosure system feed into consumer and government choices, implications

---

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

3 In a concurrent paper, Quezada-Hofflinger and Von Hippel (2017) provide complementary evidence for this channel in the case of Chile.
similar to those discussed in this paper might arise. 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).

This study relates to at least three branches of literature. First, is the literature that documents behavioral responses to high-stakes testing. These responses include diversion of resources, cheating, or manipulation of conditions under which the test is taken (see Figlio and Getzler 2002; Jacob and Levitt 2003; Jacob 2005; Figlio and Winicki 2005; Reback and Cullen 2006; Neal and Schanzenbach 2010; Apperson et al. 2016; Dee et al. 2016; Diamond and Persson 2016; Deming et al. 2016; Quezada-Hofflinger and Von Hippel 2017, among others). Behavioral responses to incentives placed by standardized testing are not, however, the only source of distortions. Mean reversion and random variation in the conditions under which the test is applied can also create distortions (see Kane and Staiger 2002; Chay et al. 2005; Graff Zivin et al. 2015; Ebenstein et al. 2016, among others). We provide evidence that non-representative test day attendance is an additional behavioral response to accountability systems and compute the implied distortions in school quality signals.

This paper also 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; Neilson 2013 and Feigenberg 2015 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). Our paper emphasizes the importance of accurate information in a context in which consumers are actively choosing.4

Finally, 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 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.

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 constructs measures of distortions in quality signals and provides an empirical discussion of their determinants. Section 5 estimates a school choice

4Our 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). We adopt it to study information frictions in 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, providing detail about the relevant institutional features.

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

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 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, across dimensions of the index, SIMCE test scores account for as much as 70 percent of the weight of the components used for its calculation (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

\footnote{The coverage of the prize was increased to 35 percent of the enrollment of the group since 2006.}
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 4th graders because they are tested every year in the period 2005–2013 and because all schools offering 4th grade also offer 1st 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). 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.6

3.1 Descriptive statistics

Using the previously described administrative records, we construct two datasets: (1) a panel of schools, and (2) 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 4th 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 4th grade. More than half of schools charge no fees, and the average monthly fee is approximately $48.7 The average 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

---

6The only measures of value added available for Chile are those computed by Neilson (2013). 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.

7All monetary units in the paper have been properly deflated and are measured in U.S. dollars using the early 2012 exchange rate.
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. We now show how attendance of students changes on test days. Because the government attempts to increase attendance on test days, and schools have incentives to prevent low-performing students taking the test, it is not a priori clear if attendance should increase or decrease on test day. Our interest is not focused on the average change in attendance, but rather in 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 4th graders ($A_{4t}$, who take the test) to the daily attendance of 3rd graders ($A_{3t}$, who do not take the test) around test days in 2013 (October 8th and 9th):

$$\Delta A = (A_{4T} - A_{3T}) - (A_{4t} - A_{3t})$$

(1)

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

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

(2)

where $A_{jkt}$ is the average attendance rate of kth graders in school $j$ and day $t$. The next section shows how a larger variance in $A_j$ translates into more distorted quality signals.

Figure 1 displays $\Delta A$ and $\Delta \overline{A}_j$. In panel (a), we plot the differential attendance rate around test days. On average, attendance increases by 2 percentage points on test days, equivalent to 0.18 standard deviations ($\sigma$). Moreover, the increase is larger among high-performing students than among low-performing students. However, despite the intuition (and anecdotal evidence) that schools might ask low-ability students to stay home on test days, we do not observe on average a decrease in attendance of students with low GPA. These averages, however, mask significant heterogeneity. In panel (b), we plot the distribution of $\Delta \overline{A}_j$. The vertical line denotes the average increase of 2 percentage points.\(^8\)

\(^8\)In order to assess the robustness of this result, we implemented the same calculations for two alternative class days: one the date of a soccer game between Chile and England and a class day exactly one week after test days. Figure A.5 displays the results for these placebo tests. In both cases, there is no differential attendance pattern across 4th and 3rd grades. Moreover, the distribution of $\Delta \overline{A}_j$ for those cases is symmetric and centered.
These attendance patterns on test days suggest that some behavioral response to the test is in place, in which pools of test-takers are not necessarily representative of their schools. Importantly, the fact that this behavior is heterogeneous across schools causes observed quality signals in the educational market to be distorted.

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. 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.\(^9\) 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.

4.1.1 Estimating missing test scores

Let us begin with the estimation of 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:

\[
q_{ijt} = f(x_{ijt}; \gamma_j) + \lambda_{jt} + \eta_{ijt}
\]  

(3)

where \( \gamma_j \) is a school specific vector of parameters, \( \lambda_{jt} \) are school-year fixed effects, and \( \eta_{ijt} \) 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 4th grade the previous year and students who studied at a different school the previous year. 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 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 around zero. This provides further support for considering the evidence presented here as a source of non-random distortions in quality signals.

\(^9\)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.
OLS.

Unsurprisingly, GPA is the strongest predictor of test scores at the student level, as displayed by Figure A.7. Moreover, given the quadratic empirical relationship between test scores and GPA, we include this variable as a quadratic polynomial. Two statistical exercises provide support for this specification. First, the R-squared of the 7,500 linear regressions we estimate are high (approximately 0.51) and are always higher in the polynomial model (see Figure A.8). 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 (Figure A.9).\footnote{A concern related to the proposed model of test scores is that of selective attendance. To test for selection, we re-estimated equation (3) 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.12. 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 for the remainder of the paper.}

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).\footnote{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.} 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.

\subsection*{4.1.2 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 $\tilde{q}_{jt}$. 

\footnote{A concern related to the proposed model of test scores is that of selective attendance. To test for selection, we re-estimated equation (3) 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.12. 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 for the remainder of the paper.}

\footnote{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.}
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.\(^{12}\)

### 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.10. Moreover, the correlation of distortions is high, as documented by Figure A.11. In what follows, we use the average of distortions across mathematics and language in 4th grade, which were taken during all years in our sample.

The average distortion is equivalent to 0.1 standard deviations of test scores at the school 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 0.06\( \sigma \) on test scores. Contreras and Rau (2012) find that the impact of SNED on test scores was between 0.14\( \sigma \) and 0.25\( \sigma \). More broadly, Kremer and Holla (2009) and Glewwe and Muralidharan (2016) review multiple educational interventions in developing countries and find impacts smaller than 0.20\( \sigma \). Similarly, a survey of field experiments in developed countries by Fryer (2016) finds that average treatment effects from school-based interventions are between 0.05\( \sigma \) and 0.07\( \sigma \). 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. In Figure A.13, we present an empirical analysis of the rank correlation between undistorted and distorted quality signals at the market-year level (see section 5.1.2 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

\(^{12}\)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. Figure A.6 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 (behavioral rather than statistical) margin that distorts quality signals.
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 \( \Delta \overline{A}_j^{\text{high}} \) and \( \Delta \overline{A}_j^{\text{low}} \)) 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? We now present a discussion of the empirical determinants of distortions. For this, we employ the panel dataset of distortions at the school level between 2005 and 2013.

4.3.1 Schools’ 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:

\[
\psi_{jmt} = X_{jt}'\theta + \nu_{mt} + \varepsilon_{jmt}
\]

where \( X_{jt} \) is a vector of school attributes in year \( t \) and \( \nu_{mt} \) 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 section 5.1.2. In order to account for the uncertainty in \( \psi_{jmt} \), 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.

Additionally, 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:

\[ \psi_{jt} = \beta X_{jt} + \eta_j + \nu_t + \varepsilon_{jt} \]  

(4)

where \( X_{jt} \) is the covariate of interest, and \( \eta_j \) and \( \nu_t \) are school and time fixed effects. Figure A.15 show basically no relationship between any of these variables and distortions.\(^{13}\)

### 4.3.2 Competitive environment

An alternative explanation is that market environment creates incentives for schools to introduce distortions and 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.\(^{14}\) 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.

\(^{13}\)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.

\(^{14}\)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 \( \eta^q \) and \( \eta^p \) 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, 2013). In our setting, we argue that observed quality \( q \) can be increased by either increasing true quality or introducing higher distortions.
4.3.3 School and teacher incentives

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

Additionally, we exploit quasi-experimental variation from two government programs to understand the role of perverse incentives. The first is the SNED teacher incentives program explained in section 2.2. This program effectively increases teachers’ wages if students in the school obtain high test scores, and it provides variation in incentives depending on the probability that a school will earn the prize (Contreras and Rau, 2012). Teachers 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. 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 win the prize, providing some evidence against the hypothesis that teachers manipulate attendance to increase test scores. See Appendix B and Figure A.18 for more detail and results.

The second government program we exploit is the ETL information program, also described in section 2.2. 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.15 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.19 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 are strongly correlated with distortions. Third, distortions are not explained

15The 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.
by within-school variation in 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 evidence against perverse incentives induced by
public programs driving variation in distortions. And finally, because individual test scores are
never disclosed, we can rule out consequences of test scores at the individual level as drivers.

Unfortunately we are not able to rule out other potential drivers of distortions. For instance,
it is possible that the existence of different school unobserved types, where some choose to dis-
tort test scores and some choose not to do so. These types might be related to, for example,
school principals, who we do not observe. Alternatively, schools’ heterogeneous responses to
idiosyncratic events might generate non-random changes in attendance on test days. For exam-
pole, schools might react to government attempts to increase attendance. While our setting does
not allow us to study these alternative explanations, that does not prevent us from estimating
the market consequences of distortions, which is what we do in the next sections.

5 Implications for school choice

This section estimates a school choice model to quantify the direct and indirect impacts of
distorted quality signals. Using the model, we implement the counterfactual exercise of pro-
viding households with undistorted quality signals and compute the direct welfare loss caused
by distortions. We emphasize heterogeneity in responses across low- and high-income students
due to differences in price- and quality-sensitivity. We then discuss policies that may help to
increase the effectiveness of the provision accurate information.

5.1 School choice model

We estimate a model of school choice in the lines of Bayer et al. (2007). 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 (2013), 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, $\xi_j$. Finally, each household has an idiosyncratic preference shock, $\varepsilon_{ij}$, which we assume is distributed T1EV.

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

$$u_{ij}^r = \sum_k x_{k,j} \beta_k^r + \xi_j^r + \beta_d^r d_{ij} + \varepsilon_{ij}$$  \hspace{1cm} (5)$$

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 (5) as follows:

$$u_{ij}^r = \delta_j^r + \beta_d^r d_{ij} + \varepsilon_{ij}$$  \hspace{1cm} (6)$$

such that the parameters of the model are contained in the vector $\beta^r$, but can be alternatively represented by the vector $\delta^r$ and by $\beta_d^r$. Note that $\delta_j^r$ 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.\footnote{In the context of school choice, there is no obvious outside option. Therefore, we instead normalize $\delta_1 = 0$ within each market.} The choice probability of school $j$ by household $i$ of type $r$ predicted by the model is a function of school and household characteristics:

$$P_{ij}^r(\delta^r, d^r, \beta_d^r) = \frac{\exp(\delta_j^r + \beta_d^r d_{ij})}{\sum_{i \in \mathcal{J}_i} \exp(\delta_i^r + \beta_d^r d_{il})}$$  \hspace{1cm} (7)$$

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.

5.1.1 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 (7), 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 estimates for $\delta^r$ and $\beta^r_d$.

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

$$
\delta^r_{jmt} = \delta^r_{0,m} + \sum_k x_k,jmt \beta_k^r + \epsilon^r_{jmt}
$$

where $\delta^r_{0,m}$ is a constant term specific to each market, year, and household type; $\beta_k^r$ 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 $\epsilon^r_{jmt}$ is a mean-zero error term. Note that $\delta^r_{0,m} + \epsilon^r_{jmt}$ maps to the unobserved school characteristic $\xi^r_{jmt}$ 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 (2013) 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, as discussed in Graff Zivin et al. (2015).$^{17}$ 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.$^{18}$

---

$^{17}$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. (2015).

$^{18}$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...
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 public, and whether a school is part of the SEP program.\footnote{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.} Finally, we are able to compute the distance between households and schools using geo-referenced data on their addresses.\footnote{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.}

5.1.2 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 (2013). Distance has been shown to be a relevant determinant of school choice in the literature (Gallego and Hernando, 2009; Neilson, 2013). In our data, students’ average distance to chosen schools is 1.3 miles and the 90\textsuperscript{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.20 provides an example for the resulting market definitions, and Table A.2 displays its summary statistics.\footnote{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.}

A description of the resulting sample is displayed in Table 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 might be driven by other factors that could be persistent in time.
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.\textsuperscript{22}

5.1.3 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.21 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 4 presents results for different specifications of instrumental variables linear regressions
of the estimates of $\delta_{jmt}$ 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.\textsuperscript{23}
Overall, the model provides a good fit for observed enrollment shares, as displayed by Figure
A.22. 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 re-
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 2013).\textsuperscript{24}

\textsuperscript{22}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.

\textsuperscript{23}Table A.3 and Table A.4 display results from the first stage of the IV estimation for school fees and quality
respectively. The bottom rows in Table 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.5 displays the results from estimating
the second stage of the model by OLS. As expected, the OLS estimates are smaller than the IV ones.

\textsuperscript{24}As a robustness check on the results, we study the correlation in estimates of unobserved school character-

19
5.2 Counterfactual analysis

In our setting, schools quality signals are distorted and therefore households are choosing schools on the basis of a misperceived vector of attributes. A key aspect, however, is that while perceived school quality might be different than true quality, the value that households ultimately obtain from a school is the true quality of their school choice. This is related to the distinction 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.

In order to compute the effects of distorted quality signals on choices and welfare, we define two scenarios: baseline and counterfactual. The former corresponds to an environment in which households actually choose schools. The latter corresponds to a counterfactual world 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 remarkably small equilibrium responses by schools.

Throughout this section, we utilize our estimates for $\delta^r$ and $\beta^r_d$, 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 additionally use estimates of $\beta^r_k$ from the second stage of the school choice model, and a counterfactual vector of school characteristics $\tilde{X}_{ij} = [p_j, \tilde{q}_j, W_j]$, where $\tilde{q}_j$ stands for the undistorted quality of school $j$.\footnote{More precisely, we utilize the results for the second stage from our preferred specifications: columns 6 and 9 of Table 4.}

5.2.1 Choices

We begin the analysis by examining school choice probabilities by households across both scenarios. We do so by adjusting the choice probabilities predicted by equation (7) of our school choice model and using parameter estimates and data on school attributes for both scenarios. Following equation (7), choice probabilities are therefore computed as $P^r_{ijmt}(d^r, \hat{\delta}^r, \hat{\beta}^r_d)$ and $P^r_{ijmt}(d^r, \tilde{\delta}^r, \tilde{\beta}^r_d)$, where $\tilde{\delta}_{jmt} = \sum_k \tilde{x}_{k,jmt} \hat{\beta}^r_k + \tilde{\xi}_{jmt}$ is the mean utility of school $j$ in market $m$ in period $t$, computed using preferences estimates and data on counterfactual school quality.

Figure A.23 displays the computed changes in choice probabilities between both scenarios. It is interesting to note that, despite the fact that the magnitude of estimated distortions is moderate, there is significant heterogeneity. This pattern holds when restricting the analysis
to the set of schools actually chosen by parents as displayed by Figures A.23-C and A.23-D. 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. Those marginal changes in the observed vector of school quality might not be strong enough as to induce households to actually change their school choices. 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.26

We compute the predicted attributes of schools chosen by households under both scenarios. Table 5 displays results for both low- and high-income households. We report both the average across switchers and across all households in the data. Columns 1 and 3 in Table 5 display results for switchers within these household groups. 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 that had 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-income (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 0.2σ (0.49σ) for low-income (high-income) switchers and, similarly, an increase in distance travelled to chosen schools of 0.04σ (0.05σ). These results imply that the subpopulation of switchers would change their choices in a substantial way. Switchers move towards higher-quality schools, for which they are willing to both travel more and pay higher fees.

Columns 2 and 4 in Table 5 display results for the average across all households. It is easy to note that changes in predicted distance to chosen schools and fees are small. This is expected since non-switching households are unaffected by the information 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.

26We calculate switching rates by simulating 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.
5.2.2 Welfare analysis

We now calculate the effects of providing undistorted quality signals on consumer surplus. In the baseline scenario households choose schools using the observed measure of 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 households choose and experience utility 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. In our setting, these two utilities are related. 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 $\tau_j$ measures the wedge between choice and experienced utility from school $j$. Under the utility function assumed in section 5.1, we know that $\tau_j = \beta_q(q_j - q_j)$.

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

$$j^*_i = \arg \max_j \{u_{ij}\}_{j \in J_i}$$

$$\tilde{j}^*_i = \arg \max_j \{\tilde{u}_{ij}\}_{j \in J_i}$$

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, $\tilde{u}_{ij^*} - \tilde{u}_{ij^*}$. Using the fact that $\tilde{u}_{ij^*} = u_{ij^*} + \tau_j$, we can compute the expected monthly change in consumer surplus as:

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

where we define $\tilde{v}_{ij} \equiv \delta_j + \beta_d d_{ij}$ 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 probabilities. Dividing by $\beta_p$ simply transforms the welfare loss to monetary units. Equation (9) calculates the average gain.

---

27 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 $\tau_j$, 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.
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 6 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, while switchers will be strictly better off. The average yearly welfare gain for switchers is $53 among low-income households and of $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.\(^{28}\)

### 5.2.3 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 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 6. 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.\(^{29}\)

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,\(^{28}\)

---

\(^{28}\)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.

\(^{29}\)Recall that in conditional logit models, coefficients are normalized by the scale parameter of the idiosyncratic preference shock, \(\sigma_\epsilon\), 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.
more than three times those in the first counterfactual and higher than those for high-income
switchers. 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.

5.2.4 Discussion

We have estimated a school choice model and studied a counterfactual exercise by which in-
formation 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
would change in the counterfactual scenario. Second, households would react to the change in
the quality disclosure system mostly by increasing the probability of choosing higher quality
schools. There would thus be 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 could increase low-income households
quality-sensitivity might increase welfare gains from this policy.

Throughout this section, we have assumed that households are uninformed about distortions
in quality signals. If they were informed, they would optimally incorporate that information
and adjust their choices according to true school quality. 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.
Theoretically, 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 already suggests that it might be socially ben-

---

30The 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 (9).
eficial to invest in quality disclosure systems that reduce distortions in educational markets. Note, however, that our counterfactual policy does not evaluate the welfare effects of the disclosure system in place, but rather the welfare effects of distorted quality signals given the current school quality disclosure system. Moreover, note that 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:

\[ I_{jgt}(q_j, q_{j-1}, X_j) = \omega_1 q_j + \omega_2 (q_j - q_{j-1}) + \omega_3 X_j \]  

(10)

where \( I_{jgt} \) is the index of school \( j \), in group \( g \), and year \( t \); \( q_j \) is the average test score in year \( t \); \( X_j \) is a vector of attributes; and \((\omega_1, \omega_2, \omega_3)\) 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 \((\tilde{q}_j, \tilde{q}_{j-1})\), we calculate schools undistorted indices using equation (10), \( \tilde{I}_{jgt} = I_{jgt}(\tilde{q}_j, \tilde{q}_{j-1}, X_j) \), 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 School choice information

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 4th and 8th 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:

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

(11)

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

Equation (11) 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 1st 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. Our results are 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. Quantifying the costs associated to other hidden actions is also necessary to fully characterize the costs of accountability systems. We highlight that the institutional environment might determine the magnitude and impacts of distortions.\textsuperscript{31} 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

\textsuperscript{31}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.6 and A.7). This suggests that our findings might be exacerbated in settings with different levels of corruption.
variables prone to distortions. Multiple programs in different countries and sectors are assigned through such rules and might be subject to misallocation.

References


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.

(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 75th percentile and below the 25th 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

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

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

**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)**

<table>
<thead>
<tr>
<th>A – School attributes</th>
<th>All</th>
<th>Distortions &gt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) (2) (3) (4)</td>
<td>(3) (4)</td>
</tr>
<tr>
<td>Public</td>
<td>1.55*** 1.31*** 0.70** 0.57*</td>
<td>0.70** 0.57*</td>
</tr>
<tr>
<td></td>
<td>(0.13) (0.14) (0.31) (0.33)</td>
<td>(0.31) (0.33)</td>
</tr>
<tr>
<td>Religious</td>
<td>0.03 -0.11 0.17 0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.08) (0.08) (0.17) (0.20)</td>
<td></td>
</tr>
<tr>
<td>For-profit</td>
<td>0.28** 0.36*** 0.76** 0.91***</td>
<td>0.76** 0.91***</td>
</tr>
<tr>
<td></td>
<td>(0.11) (0.11) (0.31) (0.32)</td>
<td>(0.31) (0.32)</td>
</tr>
<tr>
<td>Log parents income</td>
<td>-0.78*** -0.68*** -0.82*** -0.85***</td>
<td>-0.82*** -0.85***</td>
</tr>
<tr>
<td></td>
<td>(0.04) (0.04) (0.11) (0.12)</td>
<td>(0.11) (0.12)</td>
</tr>
<tr>
<td>Average annual attendance</td>
<td>-0.17*** -0.20*** -0.32*** -0.30**</td>
<td>-0.32*** -0.30**</td>
</tr>
<tr>
<td></td>
<td>(0.04) (0.05) (0.09) (0.12)</td>
<td>(0.09) (0.12)</td>
</tr>
<tr>
<td>Students in 4th grade</td>
<td>-0.12 -0.11 -2.30*** -2.22***</td>
<td>-2.30*** -2.22***</td>
</tr>
<tr>
<td></td>
<td>(0.16) (0.15) (0.25) (0.26)</td>
<td>(0.25) (0.26)</td>
</tr>
<tr>
<td>Enrollment in grades 1st-8th</td>
<td>-0.51*** -0.51*** -0.11 0.15</td>
<td>-0.11 0.15</td>
</tr>
<tr>
<td></td>
<td>(0.16) (0.16) (0.25) (0.27)</td>
<td>(0.25) (0.27)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.77*** 1.95*** 6.25*** 6.34***</td>
<td>6.25*** 6.34***</td>
</tr>
<tr>
<td></td>
<td>(0.12) (0.13) (0.31) (0.33)</td>
<td>(0.31) (0.33)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B – Autocorrelation</th>
<th>All</th>
<th>Distortions &gt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) (2) (3) (4)</td>
<td>(3) (4)</td>
</tr>
<tr>
<td>Lagged distortion</td>
<td>0.41*** 0.38*** 0.39*** 0.37***</td>
<td>0.39*** 0.37***</td>
</tr>
<tr>
<td></td>
<td>(0.01) (0.02) (0.03) (0.03)</td>
<td>(0.03) (0.03)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.97*** 2.06*** 6.25*** 6.30***</td>
<td>6.25*** 6.30***</td>
</tr>
<tr>
<td></td>
<td>(0.04) (0.05) (0.13) (0.14)</td>
<td>(0.13) (0.14)</td>
</tr>
</tbody>
</table>

| Mean of dep. variable | 2.18 2.18 5.11 5.11 | 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: Summary statistics for estimation of school choice model

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

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

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td></td>
<td></td>
<td></td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
<td>(8)</td>
<td>(9)</td>
</tr>
<tr>
<td>Fee</td>
<td>-0.003***</td>
<td>-0.004***</td>
<td>-0.006***</td>
<td>-0.006***</td>
<td>-0.007***</td>
<td>-0.010***</td>
<td>-0.002***</td>
<td>-0.003***</td>
<td>-0.005***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.001)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.000)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Quality</td>
<td>0.012***</td>
<td>0.019***</td>
<td>0.019***</td>
<td>0.004**</td>
<td>0.011***</td>
<td>0.015***</td>
<td>0.021***</td>
<td>0.028***</td>
<td>0.023***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Religious</td>
<td>-0.054**</td>
<td></td>
<td></td>
<td>-0.086***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.030)</td>
<td>(0.029)</td>
<td>(0.030)</td>
<td>(0.029)</td>
<td>(0.029)</td>
<td>(0.029)</td>
<td>(0.029)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>Gender constraint</td>
<td>0.148***</td>
<td></td>
<td></td>
<td>0.121**</td>
<td></td>
<td></td>
<td>0.161***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.059)</td>
<td>(0.055)</td>
<td>(0.059)</td>
<td>(0.059)</td>
<td>(0.055)</td>
<td>(0.055)</td>
<td>(0.055)</td>
<td>(0.055)</td>
</tr>
<tr>
<td>Public</td>
<td>0.089***</td>
<td></td>
<td></td>
<td>0.229***</td>
<td></td>
<td></td>
<td>-0.061</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.040)</td>
<td>(0.039)</td>
<td>(0.040)</td>
<td>(0.039)</td>
<td>(0.039)</td>
<td>(0.039)</td>
<td>(0.039)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>SEP school</td>
<td>-0.325***</td>
<td></td>
<td></td>
<td>-0.533***</td>
<td></td>
<td></td>
<td>-0.587***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.094)</td>
<td>(0.066)</td>
<td>(0.094)</td>
<td>(0.066)</td>
<td>(0.066)</td>
<td>(0.066)</td>
<td>(0.066)</td>
<td>(0.066)</td>
</tr>
</tbody>
</table>

| 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**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-test fee</td>
<td>1566.15</td>
<td>2031.73</td>
<td>395.15</td>
<td>484.55</td>
<td>582.07</td>
<td>73.91</td>
<td>1285.81</td>
<td>1593.35</td>
<td>329.15</td>
</tr>
<tr>
<td>F-test quality</td>
<td>70.62</td>
<td>17.69</td>
<td>15.03</td>
<td>33.81</td>
<td>9.68</td>
<td>8.17</td>
<td>36.50</td>
<td>8.03</td>
<td>6.86</td>
</tr>
<tr>
<td>Cragg-Donald EV</td>
<td>283.97</td>
<td>232.38</td>
<td>203.91</td>
<td>146.05</td>
<td>127.12</td>
<td>101.57</td>
<td>139.36</td>
<td>106.47</td>
<td>98.52</td>
</tr>
</tbody>
</table>

**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 5: Means of predicted school attributes of households choices

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Scenario</th>
<th>(1) Low-income students</th>
<th>(2) High-income students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Switchers</td>
<td>Average</td>
</tr>
<tr>
<td>Distance (in km)</td>
<td>Baseline</td>
<td>2.00</td>
<td>2.36</td>
</tr>
<tr>
<td></td>
<td>Counterfactual</td>
<td>2.07</td>
<td>2.36</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>0.07</td>
<td>0.00</td>
</tr>
<tr>
<td>Fee (in USD)</td>
<td>Baseline</td>
<td>6.58</td>
<td>17.08</td>
</tr>
<tr>
<td></td>
<td>Counterfactual</td>
<td>22.52</td>
<td>17.52</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>15.95</td>
<td>0.43</td>
</tr>
<tr>
<td>Quality (in points)</td>
<td>Baseline</td>
<td>242.72</td>
<td>254.77</td>
</tr>
<tr>
<td></td>
<td>Counterfactual</td>
<td>260.15</td>
<td>255.25</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>17.43</td>
<td>0.48</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Comparison</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(6)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counterfactual scenario</td>
<td>Switch rate</td>
<td>$53.2</td>
<td>$1.73</td>
<td>3.25%</td>
<td>$53.2</td>
<td>$1.73</td>
</tr>
<tr>
<td>Low-income households with high-income quality preferences</td>
<td>4.10%</td>
<td>$100.94</td>
<td>$4.13</td>
<td>3.04%</td>
<td>$173.94</td>
<td>$5.29</td>
</tr>
<tr>
<td>Low-income households with high-income preferences</td>
<td>3.81%</td>
<td>$181.37</td>
<td>$6.91</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Low-income households with high-income market opportunities</td>
<td>2.39%</td>
<td>$64.95</td>
<td>$1.55</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

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

Distorted Quality Signals in School Markets

José Ignacio Cuesta, Felipe González, and Cristián Larroulet

Contents

A More about statistical procedures
   A.1 Construction of bounds in multiple imputation method

B Understanding distortions in quality signals
   B.1 Monetary incentives for teachers
   B.2 Information for school choice

List of Figures

A.1 Evolution of vouchers
A.2 Timeline of standardized test scores
A.3 Test scores as quality signals
A.4 Test scores as quality signals
A.5 Comparison of school absenteeism on test day with two other events
A.6 Distortions and noise
A.7 Predictability of test scores
A.8 Prediction model
A.9 Evaluation of prediction model
A.10 Distribution of distortions by subject in 4th grade
A.11 Correlation between distortions in different tests
A.12 Heckman corrected distortions
A.13 Distribution of rank correlations over time
A.14 Distribution of rank correlations by market type
A.15 Distortions and school attributes
A.16 Distortions and attributes of schools within 3km
A.17 Potential gains and distortions ............................................. 21
A.18 Monetary incentives for teachers ............................................. 22
A.19 “Educational Traffic lights” policy ............................................. 23
A.20 Market definition ................................................................. 24
A.21 Estimated coefficients on distance from the first stage .................. 25
A.22 Observed and predicted school enrollment ................................ 26
A.23 Changes in choice probabilities .............................................. 27

List of Tables

A.1 Descriptive statistics for distortions by subject ............................. 28
A.2 School markets as connected components .................................... 29
A.3 IV results from the second stage of school choice model - First stage for school fees 30
A.4 IV results from the second stage of school choice model - First stage for school quality 31
A.5 OLS results from the second stage of school choice model ............... 32
A.6 Schools in corrupt municipalities have larger distortions .................... 33
A.7 Differences-in-differences of audits ............................................. 34
A More about statistical procedures

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:

\[
\hat{q}_{ijt} = x_{ijt}^{'} \hat{\gamma}_{jt}
\]

where \( \hat{\gamma}_{jt} \) 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 \( \hat{\gamma}_{jt} \) from the distribution \( N(\hat{\gamma}_{jt}, \hat{\Sigma}_{jt}) \), where \( \hat{\Sigma}_{jt} \) is the estimated variance-covariance matrix for \( \hat{\gamma}_{jt} \). 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.

B Understanding distortions in quality signals

B.1 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:

\[
\psi_{jt} = 1_{IN} f^{IN}(\text{SNED}^{IN}_{jt-1}) + 1_{OUT} f^{OUT}(\text{SNED}^{OUT}_{jt-1}) + \eta_j + \nu_t + \varepsilon_{jt}
\]

where \( \text{SNED}^{IN}_{jt-1} \) measures distance to the threshold for winners, and \( \text{SNED}^{OUT}_{jt-1} \) 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.18 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.2 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.

Following the discontinuous incentives at the threshold, we estimate:

\[ \psi_{jt} = 1_r f^r(q_{jt-1}) + 1_y f^y(q_{jt-1}) + 1_g f^g(q_{jt-1}) + X_{jt} + \varepsilon_{jt} \]

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.19 presents the linear relationships between test scores and distortions around the ETL policy cutoffs. Again, we present results for distortions and residualized distortions.\(^1\) 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.

References


\(^1\)Note that this is a cross-sectional exercise, so we cannot include school and year fixed effects in this case, just school characteristics.
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 2nd and 6th grade students since 2012, but we omit them from our analysis because they are relatively new.
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 as 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: Test scores as quality signals

Notes: This figure displays the relationship between test scores and the only available measure of value added in Chile, from Neilson (2013). We thank the author for providing us with this figure.
Figure A.5: Comparison of school absenteeism on test day with two other events

(a) One week after test day

(b) Soccer match day (Chile vs. England)

Notes: Panels (a) and (b) present the difference in absenteeism rates between 4th and 3rd graders across the GPA distribution around two events: (a) one week after test day, and (b) a soccer match day. The histograms in panels (c) and (d) presents 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. A Kolmogorov-Smirnov test rejects the equality of distributions in both cases ($p$-values $< 0.01$).
Figure A.6: 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.7: 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.
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.9: Evaluation of prediction model

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 25th (75th) percentile of the GPA distribution within a school-year.
Figure A.10: Distribution of distortions by subject in 4th 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.11: Correlation between distortions in different tests

Notes: These figures display the relationships between estimated distortions in test scores for different subjects of SIMCE.
Figure A.12: 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.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 $\rho_{mt}$ 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

Notes: Let $\rho_{mt}$ 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 SEP students
(g) Distorted test score  (h) Undistorted test score  (i) Students in 4th 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

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:

\[ \eta_{jmt} = \frac{\partial s_{jmt}}{\partial q_{jmt}} \frac{q_{jmt}}{s_{jmt}} = \left( \sum_r \pi_{mt}^r \frac{1}{N_{mt}^r} \sum_{i \in T_{mt}^r} \frac{\partial P_{ijmt}(d^r, \hat{\delta}_{d}, \hat{\gamma})}{\partial q_{jmt}} \right) \frac{q_{jmt}}{s_{jmt}} \]

where \( \pi_{mt}^r \) is the share of households of type \( r \) in market \( m \) and year \( t \), while \( N_{mt}^r \) and \( T_{mt}^r \) 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: Monetary incentives for teachers

Notes: Regression kink design 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.19: “Educational Traffic lights” policy

Notes: Regression kink design 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.20: 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.21: Estimated coefficients on distance from the first stage

Notes: These figures display resulting estimates for $\beta_d^*$ 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.22:** Observed and predicted school enrollment

![Graph showing observed and predicted school enrollment shares]

**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:

\[
s_{jmt}^{\hat{\delta}, \hat{\beta}_d} = \sum_{r} \pi_{rmt} \frac{1}{N_{rmt}} \sum_{i \in I_{rmt}} P_{ijmt}^{r}(d^{r}, \hat{\delta}^{r}, \hat{\beta}_d^{r})
\]

where \( \pi_{rmt} \) is the share of households of type \( r \) in market \( m \) and year \( t \), while \( N_{rmt} \) and \( I_{rmt} \) 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.23: Changes in choice probabilities

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

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

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

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

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 $\kappa$ 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 $\kappa = 5$ (highlighted in gray), but results are robust to different definitions.
Table A.3: IV results from the second stage of school choice model - First stage for school fees

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rivals’ characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religious</td>
<td>-19.876**</td>
<td>79.768*</td>
<td>100.580</td>
<td>-11.500</td>
<td>68.481</td>
<td>11.209</td>
<td>-26.668**</td>
<td>85.263</td>
<td>195.296*</td>
</tr>
<tr>
<td>Gender constraint</td>
<td>-80.777***</td>
<td>463.861***</td>
<td>780.010***</td>
<td>-90.315***</td>
<td>513.487***</td>
<td>617.324**</td>
<td>-62.196**</td>
<td>418.205**</td>
<td>922.088***</td>
</tr>
<tr>
<td></td>
<td>(6.075)</td>
<td>(45.861)</td>
<td>(61.668)</td>
<td>(7.921)</td>
<td>(74.814)</td>
<td>(94.836)</td>
<td>(9.017)</td>
<td>(74.814)</td>
<td>(94.836)</td>
</tr>
<tr>
<td><strong>Teacher wages</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average hourly wage</td>
<td>0.094***</td>
<td>0.123***</td>
<td>0.135***</td>
<td>0.060</td>
<td>0.078*</td>
<td>0.089**</td>
<td>0.131***</td>
<td>0.170***</td>
<td>0.184***</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.032)</td>
<td>(0.030)</td>
<td>(0.042)</td>
<td>(0.043)</td>
<td>(0.041)</td>
<td>(0.045)</td>
<td>(0.047)</td>
<td>(0.045)</td>
</tr>
<tr>
<td><strong>Voucher</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>-1.335***</td>
<td>-1.473***</td>
<td>-1.454***</td>
<td>-0.975***</td>
<td>-1.177***</td>
<td>-1.166***</td>
<td>-1.535***</td>
<td>-1.627***</td>
<td>-1.603***</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.030)</td>
<td>(0.029)</td>
<td>(0.059)</td>
<td>(0.055)</td>
<td>(0.054)</td>
<td>(0.037)</td>
<td>(0.036)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>SEP school</td>
<td>0.062*</td>
<td>-0.080**</td>
<td>0.007</td>
<td>-0.093***</td>
<td>-0.217***</td>
<td>-0.077**</td>
<td>0.210***</td>
<td>0.096</td>
<td>0.090*</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.033)</td>
<td>(0.032)</td>
<td>(0.034)</td>
<td>(0.035)</td>
<td>(0.033)</td>
<td>(0.052)</td>
<td>(0.052)</td>
<td>(0.049)</td>
</tr>
<tr>
<td></td>
<td>(0.185)</td>
<td>(0.187)</td>
<td>(0.546)</td>
<td>(0.235)</td>
<td>(0.241)</td>
<td>(0.636)</td>
<td>(0.279)</td>
<td>(0.279)</td>
<td>(0.815)</td>
</tr>
<tr>
<td><strong>Temperature</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>4.167***</td>
<td>2.684</td>
<td>0.621</td>
<td>3.980**</td>
<td>3.871</td>
<td>0.808</td>
<td>4.284**</td>
<td>0.709</td>
<td>-0.311</td>
</tr>
<tr>
<td></td>
<td>(1.308)</td>
<td>(4.491)</td>
<td>(4.381)</td>
<td>(1.778)</td>
<td>(6.120)</td>
<td>(5.784)</td>
<td>(1.874)</td>
<td>(6.490)</td>
<td>(6.466)</td>
</tr>
<tr>
<td>Quadratic</td>
<td>-0.261***</td>
<td>-0.163</td>
<td>-0.095</td>
<td>-0.256***</td>
<td>-0.214</td>
<td>-0.110</td>
<td>-0.260***</td>
<td>-0.088</td>
<td>-0.058</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.179)</td>
<td>(0.174)</td>
<td>(0.074)</td>
<td>(0.243)</td>
<td>(0.230)</td>
<td>(0.078)</td>
<td>(0.259)</td>
<td>(0.258)</td>
</tr>
<tr>
<td><strong>SNED program</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.211)</td>
<td>(1.159)</td>
<td>(1.433)</td>
<td>(1.560)</td>
<td>(1.499)</td>
<td>(1.472)</td>
<td>(1.816)</td>
<td>(1.727)</td>
<td>(1.698)</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Market-Year F.E.</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>10,774</td>
<td>10,774</td>
<td>10,774</td>
<td>5,335</td>
<td>5,335</td>
<td>5,335</td>
<td>5,439</td>
<td>5,439</td>
<td>5,439</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.685</td>
<td>0.729</td>
<td>0.739</td>
<td>0.665</td>
<td>0.716</td>
<td>0.730</td>
<td>0.706</td>
<td>0.748</td>
<td>0.758</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(4) All Low-income students</td>
<td>(4) High-income students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rivals’ characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religious</td>
<td>1.173**</td>
<td>12.207**</td>
<td>62.650**</td>
<td>2.547**</td>
<td>9.472**</td>
<td>54.491**</td>
<td>-0.046**</td>
<td>14.997*</td>
<td>71.121*</td>
</tr>
<tr>
<td></td>
<td>(3.687)</td>
<td>(20.849)</td>
<td>(29.730)</td>
<td>(5.269)</td>
<td>(29.957)</td>
<td>(42.769)</td>
<td>(5.157)</td>
<td>(29.337)</td>
<td>(41.727)</td>
</tr>
<tr>
<td></td>
<td>(2.729)</td>
<td>(22.419)</td>
<td>(30.724)</td>
<td>(3.881)</td>
<td>(32.046)</td>
<td>(43.895)</td>
<td>(3.834)</td>
<td>(31.777)</td>
<td>(43.497)</td>
</tr>
<tr>
<td><strong>Teacher wages</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average hourly wage</td>
<td>0.060***</td>
<td>0.064***</td>
<td>0.072***</td>
<td>0.059***</td>
<td>0.062***</td>
<td>0.071***</td>
<td>0.061***</td>
<td>0.067***</td>
<td>0.074***</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(0.013)</td>
<td>(0.018)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.018)</td>
<td>(0.019)</td>
<td>(0.019)</td>
</tr>
<tr>
<td><strong>Voucher</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>-0.212***</td>
<td>-0.247***</td>
<td>-0.234***</td>
<td>-0.183***</td>
<td>-0.231***</td>
<td>-0.219***</td>
<td>-0.227***</td>
<td>-0.254***</td>
<td>-0.241***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.016)</td>
<td>(0.016)</td>
<td>(0.015)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>SEP school</td>
<td>0.715***</td>
<td>0.604***</td>
<td>0.625***</td>
<td>0.720***</td>
<td>0.615***</td>
<td>0.636***</td>
<td>0.709***</td>
<td>0.593***</td>
<td>0.614***</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.017)</td>
<td>(0.017)</td>
<td>(0.025)</td>
<td>(0.024)</td>
<td>(0.024)</td>
<td>(0.025)</td>
<td>(0.025)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>SEP share</td>
<td>-4.665***</td>
<td>-4.149***</td>
<td>-0.157</td>
<td>-4.694***</td>
<td>-4.209***</td>
<td>-0.177</td>
<td>-4.628***</td>
<td>-4.092***</td>
<td>-0.134</td>
</tr>
<tr>
<td></td>
<td>(0.082)</td>
<td>(0.079)</td>
<td>(0.202)</td>
<td>(0.115)</td>
<td>(0.113)</td>
<td>(0.289)</td>
<td>(0.117)</td>
<td>(0.113)</td>
<td>(0.286)</td>
</tr>
<tr>
<td><strong>Temperature</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>4.151***</td>
<td>1.492</td>
<td>0.620</td>
<td>4.134***</td>
<td>1.727</td>
<td>0.620</td>
<td>4.166***</td>
<td>1.196</td>
<td>0.554</td>
</tr>
<tr>
<td></td>
<td>(0.494)</td>
<td>(1.630)</td>
<td>(1.489)</td>
<td>(0.712)</td>
<td>(2.339)</td>
<td>(2.115)</td>
<td>(0.686)</td>
<td>(2.298)</td>
<td>(2.122)</td>
</tr>
<tr>
<td>Quadratic</td>
<td>-0.208***</td>
<td>-0.091</td>
<td>-0.064</td>
<td>-0.206***</td>
<td>-0.105</td>
<td>-0.067</td>
<td>-0.209***</td>
<td>-0.076</td>
<td>-0.059</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.066)</td>
<td>(0.060)</td>
<td>(0.030)</td>
<td>(0.094)</td>
<td>(0.085)</td>
<td>(0.029)</td>
<td>(0.093)</td>
<td>(0.085)</td>
</tr>
<tr>
<td><strong>SNED program</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prize residual</td>
<td>1.713***</td>
<td>3.029***</td>
<td>2.412***</td>
<td>1.861***</td>
<td>3.143***</td>
<td>2.492***</td>
<td>1.598***</td>
<td>2.935***</td>
<td>2.349***</td>
</tr>
<tr>
<td></td>
<td>(0.486)</td>
<td>(0.447)</td>
<td>(0.431)</td>
<td>(0.687)</td>
<td>(0.636)</td>
<td>(0.614)</td>
<td>(0.689)</td>
<td>(0.634)</td>
<td>(0.612)</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Market-Year F.E.</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>10.774</td>
<td>10.774</td>
<td>10.774</td>
<td>5.335</td>
<td>5.335</td>
<td>5.335</td>
<td>5.439</td>
<td>5.439</td>
<td>5.439</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.391</td>
<td>0.474</td>
<td>0.512</td>
<td>0.367</td>
<td>0.452</td>
<td>0.492</td>
<td>0.412</td>
<td>0.493</td>
<td>0.529</td>
</tr>
</tbody>
</table>

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.5: OLS results from the second stage of school choice model

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

**Notes:** All regressions are weighted by school enrollment. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Table A.6: Schools in corrupt municipalities have larger distortions

Dependent variable is distortions (in test score points)

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

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.7: Differences-in-differences of audits

*Dependent variable is distortions (in test score points)*

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

*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.