Journal of Development Economics 147 (2020) 102532

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular Article

Distorted quality signals in school markets☆

José Ignacio Cuesta [a,*], Felipe González [b], Cristian Larroulet Philippi [c]

[a] Stanford University, USA
[b] Pontificia Universidad Católica de Chile, Chile
[c] University of Cambridge, UK

ARTICLE INFO

JEL classification:
I20
L15

Keywords:
Accountability
Schools
Quality
Disclosure
Choice

ABSTRACT

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

1. Introduction

Information plays a key role in markets with consumer choice and government intervention. In educational markets, standardized tests have been used for decades in developed countries as a common practice for collecting information about schools and students for the purposes of consumer choice and government interventions (Figlio and Loeb, 2011). Developing countries have followed a similar trend, as increased expenditure and enrollment have motivated accountability systems based on test scores as a means to increase school quality (Mbiti, 2016). However, accountability systems that resort to these tests have been controversial among academics and educators. Critics argue that, because of undesirable behavioral responses and/or measurement problems that distort actual test scores, the usage of standardized tests in accountability systems is problematic (e.g., Figlio and Getzler, 2002; Kane and Staiger, 2002; Neal, 2013). How large are distortions in test scores? And, more importantly, what are the market and policy consequences of these distortions? Despite increasing evidence of undesirable behavioral responses and measurement problems, quantification of these distortions and their consequences is surprisingly lacking.

In this paper, we quantify distortions in school quality signals and their market consequences. We study one of the most developed accountability systems in the world—Chile's market-oriented educational system (Figlio and Loeb, 2011). The Chilean government relies on standardized testing to generate school-specific quality signals. These measures are used not only for quality assessment and performance evaluation, which are key inputs in the policy process, but also as a disclosure system in school choice. These features make Chile an ideal setting to quantify the consequences of distortions in test scores on both household school choices and the allocation of public programs. We show that heterogeneous test day attendance distorts the school-specific quality metric relevant in this market. Moreover, we show that these distortions are large and have significant consequences both on school choice and on the allocation of public programs.

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

* Corresponding author.
E-mail addresses: jicuesta@stanford.edu (J.I. Cuesta), fagonza4@uc.cl (F. González), cl792@cam.ac.uk (C. Larroulet Philippi).

https://doi.org/10.1016/j.jdeveco.2020.102532
Received 23 October 2019; Received in revised form 17 July 2020; Accepted 21 July 2020
Available online 11 August 2020

The analysis proceeds in four steps. First, we show that high-performing students are more likely to attend on test days than their low-performing classmates. 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. Although school attendance in Chile is relatively high, and it increases from approximately 92 to 95 percent on test days, we find that students of high academic performance *increase* their attendance by 3 percentage points and students of low performance *decrease* it by 2 percentage points. Notably, this degree of student non-representativeness varies considerably across schools.

Second, we estimate distortions in school-level test scores using a multiple imputation method to predict the test scores of absent students.[1] We find average distortions in the system to be 0.10 standard deviations of school test scores. Importantly, distortions vary widely across schools and are persistent within schools over time. We provide support for our imputation approach using cross validation and by accounting for selective attendance. Our analysis strongly suggests that the patterns of absenteeism on test days and distortions are *not* random.

Third, to quantify the implications of these distortions we estimate a school choice model. We find that providing undistorted school quality information would likely induce three percent of students in a cohort to switch schools. To estimate the model, we use geocoded addresses of 100,000 students and 1500 schools, and estimate a discrete choice model in which, given the absence of school attendance zones, households trade-off school quality and distance. For identification, we exploit quasi-experimental variation in government programs and fixed characteristics of competitors. Given the magnitude of distortions and the spatial distribution of schools, the trade-off between distance to school and quality explains the student switching rate among schools. Our results suggest that households that would change their choices are willing to pay 117 U.S. dollars annually for undistorted quality information, with high-income households willing to pay more than low-income households due to differences in preferences over school fees and school quality.

Fourth, we show that two large public programs are 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 every year, equivalent to $20 million U.S. dollars in the last twenty years. In the second program, the government used test scores to classify schools in three quality categories and delivered this information to parents with the objective of assisting school choice. Using the classification algorithm, we show that four percent of schools were incorrectly classified and these errors persuaded two percent of the incoming student cohort to choose a different school.

This paper makes three main contributions. First, we document a novel channel through which school performance measures can get distorted: relative attendance on test days of high/low-performing students.[2] Changes in test day relative attendance of high/low-performing students have not been documented in previous work, where the most common sources of distortion are manipulation of the testing pool via selective assignment of students to special education programs (Jacob, 2005; Cullen and Reback, 2006; Lemke et al., 2006; Rockoff and Turner, 2010; Figlio and Loeb, 2011) and selective application of disciplinary measures (Figlio, 2006). Second, this is the first paper that implements a statistical method to *quantify* the magnitude of the distortions in quality signals that arise from non-representative attendance. Estimating missing test scores differs markedly from penalties such as imputing the worst score, which could harm schools with low attendance rates. Moreover, in contrast to previous work documenting noise and volatility in test scores and the associated school rankings (Kane and Staiger, 2002; Chay et al., 2005), we focus on a systematic source of error in test scores that is fixable using standard and widely used statistical techniques.

Our third contribution is that we estimate the *effects* of distortions both on school choice and the allocation of public programs. This is the first paper to quantify the marketwide consequences of distortions on educational systems. While we implement our analysis in the Chilean educational market, the implications of it go beyond both Chile and schooling. School choice has became increasingly important in educational systems in the developing world (Baum et al., 2014; Muralidharan and Sundararaman, 2015), and our findings inform the design of accountability systems for those settings. Furthermore, multiple markets in which quality is imperfectly observed have quality disclosure systems, many of which may be prone to be distorted (Dranove and Jin, 2010). Whenever quality signals generated by the disclosure system feed into consumer and government choices, implications similar to those discussed in this paper might arise.[3] In this line, we highlight that the educational system we study is well-functioning: the overall high attendance rates of Chilean students is comparable to that in the U.S. and other developed countries. Therefore, we conjecture that the type of distortions we document could be even larger in countries with low attendance rates and weak institutions, which naturally increases the value of this type of analysis.

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

This paper also contributes to the school choice literature. Previous research from developed and developing countries 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, 2017 and Sánchez, 2018 for Chile; Andrabi et al., 2017; Bau, 2019 and Carneiro et al., 2019 for Pakistan; Bayer et al., 2007; Hastings et al., 2009 and Walters, 2018 for the U.S., among others). In addition, another set of studies investigates how information affects school choice, which has found mixed results but shows that information has the potential to aid household educational decisions (Hastings and Weinstein, 2008; Jensen, 2010; Mizala and Urquiola, 2013; Andrabi et al., 2017; Allende et al., 2019). Information programs

[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).

[2] In concurrent papers, Sánchez (2019) and Quezada-Hofflinger and Von Hippel (2018) provide complementary evidence for this channel in Chile. However, their analysis is different as they focus on how test day attendance distorts the evaluation of a voucher program introduced in 2008, similar to Feigenberg et al. (2018).

[3] 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).

are among the demand-side interventions that have received the most recent attention by education research in developing country settings (Muralidharan, 2017). Our paper emphasizes the importance of *accurate* information in a context in which consumers are actively choosing.

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 a brief discussion of their determinants. Section 5 estimates a school choice model and studies the choice and welfare implications of distorted quality signals. Section 6 shows that two large public programs are misallocated because of non-random attendance on test days. Finally, section 7 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 details 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).[4]

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. Finally, both the autonomy in spending and the renewal in funding provided by this program is attached to school test scores. For further detail on the SEP program, see Correa et al. (2014).

The National System of Quality Measurement (*Sistema de Medición de la Calidad de la Educación*, SIMCE) has existed since 1988 and gives national standardized tests on different subjects. Tests are implemented every year at the national level for a subset of grades. 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.[5]

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

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

3. Data and attendance on test days

We use four administrative datasets provided by the Ministry of Education. First, is the record of schools operating between 2005 and 2013, in which we observe school type (public, private-voucher, private non-voucher), enrollment, fees, participation in government programs, and school addresses, which we use to construct markets. Second, we use student records between 2005 and 2013 (approximately 3.5 million per

[4] Fig. A.1 displays the evolution of the amount covered by vouchers during the years included in our dataset, and shows that the amount paid to schools offering JEC is larger than what other schools receive.

[5] Though the Chilean assessment system is widely considered to be of high-stakes, we emphasize that not all of the accountability mechanisms present in the U.S. context are so in Chile. In particular, during our period of study Chilean schools did not face the threat of being closed by the government, as it happens in the U.S. under No Child Left Behind (NCLB). Of course, other accountability mechanisms are present to a larger extent in Chile than in the U.S., including school choice.

Table 1
Descriptive statistics.

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

Notes: Own construction based on administrative data provided by the Ministry of Education. We restrict the data to schools with zero distortion or with sufficient data to calculate it. Distortions are measured in test score points and we estimated them using the methodology described in section 4.1. See section 4 for details. There are 8254 schools in the period 2005–2013.

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.[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 5386 different schools and, on average, 4640 schools operating in a given year. Table 1A 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 1B presents descriptive statistics for the student level dataset. Students' academic performance is measured by their GPA, which ranges from 1 to 7, with a threshold of 4 as passing grade. The mean of this variable is 5.9. The last two variables are attendance rates on test and non-test days. The former is simply the average of two indicator variables that take the value of one if a student went to school on test days; there are two test days, so this variable has the value of 0, 0.5, or 1 at the student level. The latter is the average attendance in the five non-test days previous to test ones.

3.2. Attendance on test days

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

In order to estimate the average change in students' attendance on test days, we compare the daily attendance rate of 4th graders ($\overline{A}_{4t}$, who take the test) to the daily attendance of 3*rd* graders ($\overline{A}_{3t}$, who do not take the test) around test days in 2013 (October 8th and 9th):

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

where $t = T$ represents the two test days, and $t = \tau$ represent other days around test days. We calculate $\Delta\overline{A}$ in four subsamples of students:

[6] Fig. A.2 shows how test scores are publicly disseminated through media outlets, used for advertising by schools, and used as policy tools by the government. The only measures of value added available for Chile are those computed by Neilson (2017). These value added measures are based on confidential administrative data. Fig. A.3 displays the relationship between that measure of value added and test scores, which is positive and strong.

[7] All monetary units in the paper are measured in U.S. dollars using the early 2012 exchange rate.

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 $\Delta\overline{A}$, we calculate the following school-specific changes in attendance on test day:

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

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

Fig. 1a plots the differential attendance rate around test days, $\Delta\overline{A}$. On average, attendance increases by 2 percentage points on test days, equivalent to 0.18 standard deviations (σ).[8] More interestingly, however, we find that high-performing students increase their attendance by 3 percentage points and low-performing students decrease their attendance by 2 percentage points. Although the latter decrease is consistent with anecdotal evidence, the average increase in attendance among high-performing students is somewhat surprising. Importantly, these averages mask significant heterogeneity. Panel (b) plots the distribution of $\Delta\overline{A}_j$. The vertical line denotes the average increase of 2 percentage points.[9]

Overall, these patterns suggest that heterogeneity in attendance on test days is not the result of statistical noise, but rather that some behavioral response to the accountability system is in place. There are several candidates for what drives these patterns. On the one hand, this pattern might be driven by school behavioral responses to the incentives they face. On the other hand, this pattern may simply reflect that high-performing students are more likely to attend on test days for some reason. These explanations need not be mutually exclusive. In any case, these patterns create distortions in school test scores. Importantly, the fact that this pattern is heterogeneous *across* schools causes observed quality signals in the educational market to be distorted in ways that affect school choice and the allocation of public programs.

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.[10] 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.[11] 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.

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:

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

where γ_j is a school specific vector of parameters, λ_{jt} are school-year fixed effects, and η_{ijt} is a mean zero random error term. Importantly, the vector x_{ijt} needs to contain strong predictors of test scores and be available for *all* students. We choose GPA and the following indicator variables: students who were in 4th grade the previous year and students who studied at a different school the previous year. Unsurprisingly, GPA is the strongest predictor of test scores at the student level. Moreover, given the quadratic empirical relationship between test scores and GPA, we include this variable as a quadratic polynomial, i.e. $f(a) = \tau_1 a + \tau_2 a^2$. The remaining variables are indicators and thus enter linearly. Note that equation (3) allows for the gradient of test scores to covariates in x_{ijt} to vary across schools. There are 7500 schools in our dataset with, on average, 270 test takers between 2005 and 2013. This means that our imputation method relies on 7500 regression equations that use on average 270 observations and that we estimate using OLS.

We use equation (3) to predict test scores for absent students in the period 2005–2013. In order to account for the uncertainty related to the estimation of missing test scores, we estimate these test scores multiple times by drawing from the asymptotic variance of the estimated parameters $\widehat{\gamma}_j$, an approach similar to that in Mas and Moretti (2009).[12] 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.

After estimating test scores of absent students, we estimate "undistorted" quality signals using a simple simulation estimator. Let $\widetilde{q}_{jt}^{(n)}$ 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:

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

The uncertainty of our estimates corresponds to the variance of the imputations $\widetilde{q}_{jt}^{(n)}$. We order $\widetilde{q}_{jt}^{(n)}$ 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 $\widetilde{q}_{jt}$.

We define distortions in quality signals as $\psi_{jt} \equiv q_{jt} - \widetilde{q}_{jt}$, where q_{jt} is the observed (distorted) quality signal of school j in year t. Thus, a school with a positive distortion $\psi_{jt} > 0$ is one that signaled a higher quality than its true quality through its test score, i.e. $q_{jt} > \widetilde{q}_{jt}$. Each distortion in our dataset has an associated distribution and a corresponding

[8] To put these numbers in perspective, note first that the annual average daily attendance of our sample is high at 93.3%, as shown by Table 1. Indeed, it is not different from the national U.S. annual average daily attendance of elementary schools, which was of 94% in 2007–2008 (NCES, 2011). Because NCLB requires a minimum of 95% of test-takers, one would also expect attendance on test days in the US to increase relative to non-test days. Lemke et al. (2006) show that a 6% of their sample of 11th grade students did not take any of the tests, which likely reflects a small increase in test day attendance when comparing to NCES (2011). Thus, attendance patterns in Chile might not differ much from those in the U.S. Comparisons with other developing countries are difficult to perform, due to the lack of national data on daily attendance. Still, Barrera-Osorio et al. (2011) report an average attendance of 79% among vulnerable children families in Colombia, and ASER (2019) reports an average attendance rate for rural India of around 71% in the last decade.

[9] As a placebo exercise, we implemented the same calculations for other dates and, reassuringly, we find that attendance patterns across 4th and 3*rd* grades to be similar. See Fig. A.4 for these results.

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

[11] 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.

[12] To construct bounds for distortions, we take $S = 100$ draws of $\widehat{\gamma}_{jt}$ from the distribution $N(\widehat{\gamma}_{jt}, \widehat{\Sigma}_{jt})$, where $\widehat{\Sigma}_{jt}$ is the estimated variance-covariance matrix for $\widehat{\gamma}_{jt}$. As a result, we have one hundred estimated test scores for each student that did not take the test and, by calculating the average test score for each school-year, one hundred undistorted quality signals. We construct bounds for distortions using the percentiles 2.5 and 97.5 of these one hundred undistorted signals. An alternative bootstrap procedure delivers similar results.

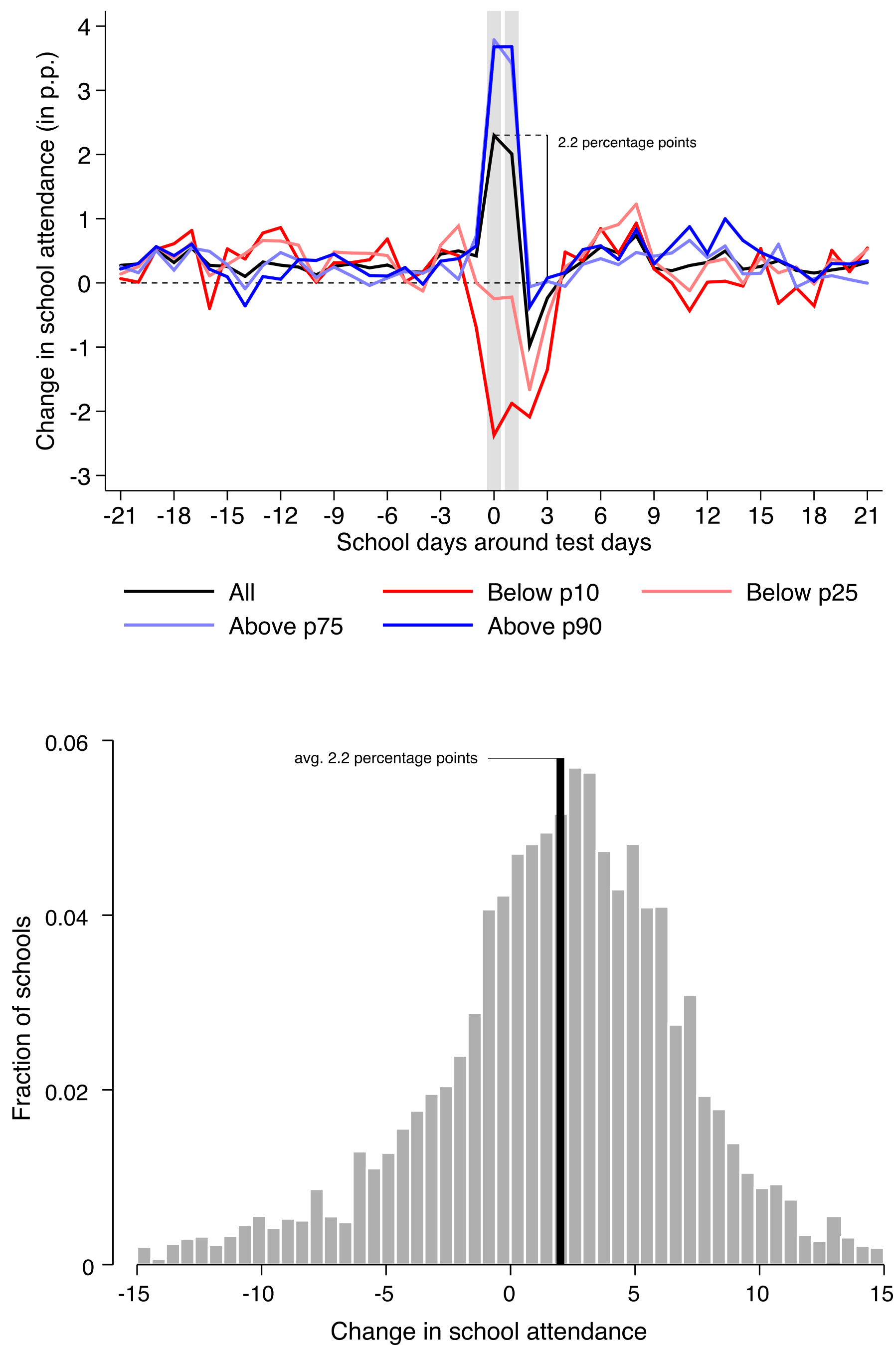

Fig. 1. School attendance around test days.
(a) Difference in average attendance rate (*y*-axis, in percentage points) between 4*th* graders (test takers) and 3*rd* graders (non-takers) around the two test days in 2013 (*x*-axis). Students are grouped by their position in the school GPA distribution. (b) Distribution of changes in school attendance in test days in 2013 (in percentage points).

confidence interval which size depends on the uncertainty associated to the model in equation (3).

4.2. Descriptive statistics of distortions

The average distortion has a value of 2.7 test score points, equivalent to 0.10 standard deviations (σ) of test scores at the school level.[13,14] There is substantial heterogeneity across schools.

Fig. 2a 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

[13] We use the average of distortions across the math and language tests in 4th grade. The Appendix presents descriptive statistics for other subjects. The distribution of distortions is remarkably similar across subjects and the correlation of distortions across subjects is high. See Figs. A.5 and A.6.

[14] To put this magnitude in context, note that the average impact of important educational policies on schools lies typically between the 0.05–0.025σ range (see Bellei, 2009; Contreras and Rau, 2012 for interventions in Chile, Kremer and Holla, 2009; Glewwe and Muralidharan, 2016 in developing countries, and Fryer, 2017, in developed ones).

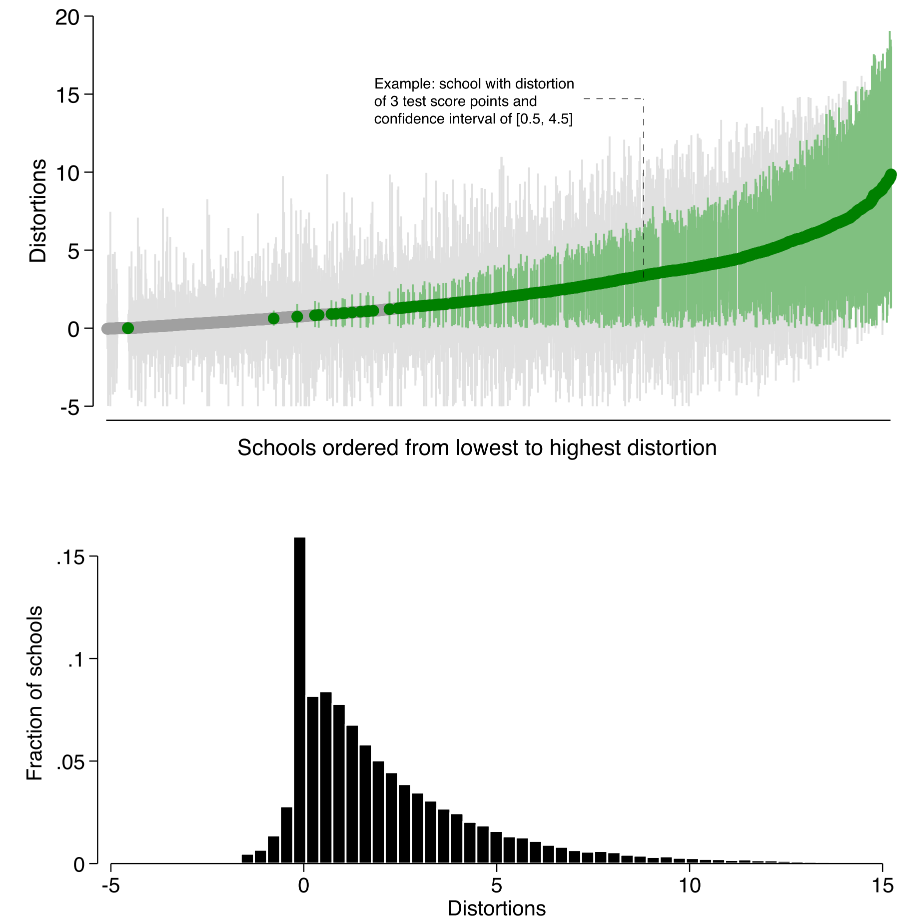

Fig. 2. Distortions in quality signals.
(a) Distortion in quality signals (y-axis, in test score points) are defined as (minus) the difference between school's observed test score and school's counterfactual test score. Schools are ordered from lower to higher distortions in the x-axis. Vertical lines 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.

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. Fig. 2b 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.

Finally, we relate the estimated distortions with the motivating evidence presented in section 3.2. We would expect schools with higher differential changes in attendance in test days for high performing students (i.e. the difference between $\Delta\overline{A}_j^{high}$ and $\Delta\overline{A}_j^{low}$) to display larger distortions in quality signals. In this line, we start by calculating the difference in $\Delta\overline{A}_j$ between students above the 75th percentile and below the 25th percentile of the school's GPA distribution. Then, we study the relationship between this measure and our estimated distortions, displayed in Fig. 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. Discussion

What explains the variation in distortions? Individual test scores are never disclosed to schools or students and therefore we can rule out incentives for students as drivers of distortions. In Appendix B, we present a discussion of the determinants of distortions, in which we focus on a variety of school level characteristics and incentives as potential drivers of them. Importantly, we observe that distortions are autocorrelated within schools, further confirming that these do not arise due to a random phenomenon.

Overall, we find that fixed characteristics of schools and some features of a competitive environment seem to be important drivers of distortions. In particular, exploiting the panel data of schools observed yearly between 2005 and 2013, we calculate that about one-third of the variance in distortions can be explained with school fixed effects. In contrast, using the same panel data and school fixed effects we observe that distortions seem not to be driven by within-school variation in relevant observable school characteristics. Distortions are larger in small public schools, for-profit schools, and schools with low attendance rates. To test for the importance of the competitive environment, we follow Dorfman and Steiner (1954) to calculate the quality demand elasticity faced by schools and also construct a measure of "potential gains" based on the theoretical absenteeism of the ten percent of students in the bottom of the GPA distribution. We find that distortions increase sharply with market competitiveness and our measure of potential gains. Finally, we find mixed evidence for how public programs relate to distortions. In particular, government funding attached to test scores displays a positive relationship with distortions, while both teacher monetary incentives and quality disclosure policies display no relationship with them. More details can be found in Appendix B.

Taken together, these results suggest that at least part of the distortions in quality signals are driven by strategic school behavioral responses to market and government incentives. Regardless of the exact factors driving distortions, we can study the consequences of distortions both on school choice and on the allocation of public programs. To motivate this exercise, recall that these distortions are on average sizable and that they vary substantially across schools. This heterogeneity matters in contexts where test scores are widely disseminated and parents can choose schools. We know from previous work that school

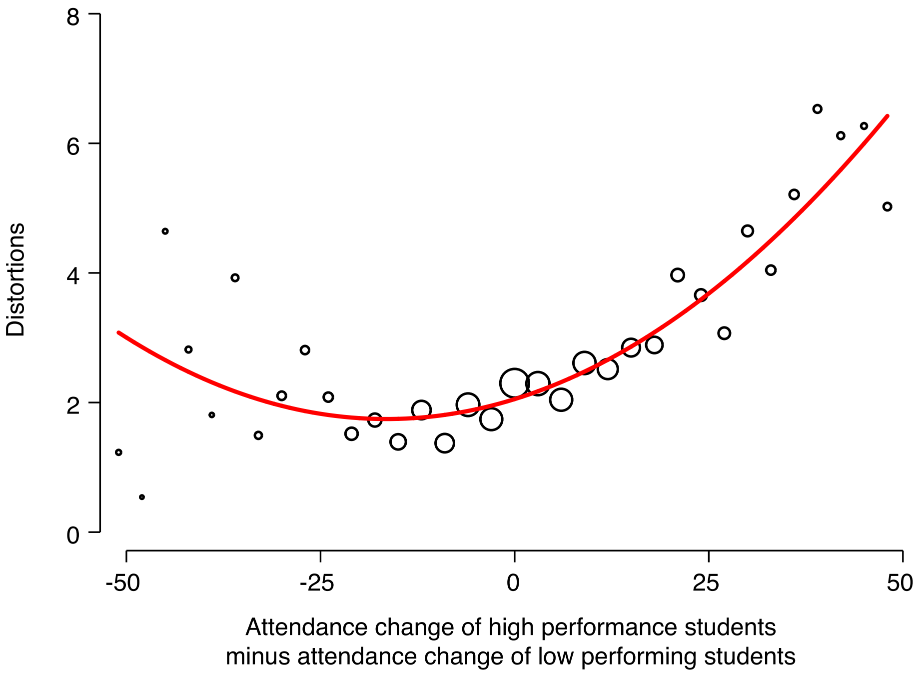

Fig. 3. Distortions and attendance in test days.
Notes: This figure displays the differential test-day attendance of students above the 75*th* percentile and below the 25*th* percentile of the GPA distribution (x-axis, in percentage points) and distortions in quality signals (y-axis, in test score points).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

choice is sensitive to quality signals (Bayer et al., 2007; Gallego and Hernando, 2009; Hastings et al., 2009; Neilson, 2017). Moreover, we also know that even small changes in test scores can affect the allocation and evaluation of public programs, because many such programs rely on school classifications heavily based on average test scores (e.g., Lemke et al., 2006; Chay et al., 2005). This motivates the remainder of the paper, in which we quantify the effects of distortions on these margins.

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, and in the context of the quality disclosure system in place in Chile, we implement a counterfactual analysis to study the effects that accounting for the distortions we estimate when constructing quality signals would have on school choice. 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 providing accurate information.

5.1. School choice model

We estimate a model of school choice in the lines of Bayer et al. (2007) and Neilson (2017). 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 (2017), this assumption is likely to hold in the Chilean school system. Finally, we assume the household's location choice is independent of the school choice problem. This assumption is supported by the lack of constraints on the choice set of schools based on residential location.

Let households be indexed by i and schools by j. Household utility depends on school fees, quality, and distance to school, denoted respectively p_j, q_j and d_{ij}. They also derive utility from other school characteristics W_j. For notational simplicity, we denote $X_j = [p_j, q_j, W_j]$, which includes K attributes. Preferences are heterogeneous depending on household type, indexed by r. In our model, only observed heterogeneity in preferences is considered, as explained below. Moreover, we allow for households to derive utility from schools' characteristics that are unobserved to the econometrician, ξ_j. Examples of unobserved school characteristics are infrastructure, special programs, among others. Finally, each household has an idiosyncratic preference shock, ε_{ij}, which we assume is distributed iid T1EV.

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

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

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

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

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

The probability of household i choosing school j can be derived analytically using households indirect utility.[15] The choice probability of school j by household i of type r predicted by the model is a function of school and household characteristics:

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

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

where $\mathcal{J}_i$ is the set of schools in the market where household i is located. We exploit this result along with data on school choices and attributes to estimate household preferences over schools.

5.1.1. Estimation

We employ a two-step procedure to estimate the parameters of the model. 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 school attributes and recover preference parameters.

The first stage of the estimation procedure consists of estimating equation (6) 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 δ^r and β^r_d.

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

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

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

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 to focus on the margin in which most school choices are made. The vector X_j includes 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.[16] Finally, we are able to compute the distance between households and schools using geo-referenced data on their addresses.[17]

5.1.2. Identification

The identification of the model relies on a combination of exogeneity assumptions and instrumental variables. The first stage parameters δ^r_{jmt} and β^r_d are identified by assuming that household location is independent of the school choice problem and that preference shocks ε_{ij} are iid. For the identification of preferences over school attributes in the second stage, a concern is the potential endogeneity of school characteristics—particularly of fees and quality, which may be correlated with unobserved school attributes ξ^r_{jmt}. To address this concern, we adopt an instrumental variables approach to estimate equation (7). Instruments must be related to school prices and quality, but unrelated to school unobservables.

We employ three sets of instruments for school fees. First, we use functions of fixed non-price and non-quality characteristics of other schools in the market. In particular, we compute the share of religious schools, schools with gender constraints, and public schools in the market among rivals for each school in the sample. These instruments resemble those in Berry et al. (1995), which are commonly used for demand estimation. Facing tougher competition affects school pricing behavior, but should be unrelated to school unobservable attributes ξ^r_{jmt}.[18] These instruments would fail if schools adjust their observable attributes in response to unobservable preference shocks, which motivates focusing on attributes that schools are unlikely to adjust in the short run. Second, we follow Neilson (2017) and employ average teacher hourly wages as an instrument. Teacher wages arguably operate as a cost shifter that affects school fees.[19] A concern for the exogeneity condition of this instrument is the potential sorting of teacher quality on the basis of school unobservable attributes ξ^r_{jmt} conditional on observables attributes X_j. Third, we use the funding amounts of different voucher programs described in section 2.2 as additional instruments. The variation in these programs is driven by policy changes and school characteristics that are fixed in the short run, such as participation in the SEP program. These instruments would fail if schools adjusted those characteristics in response to unobservable preference shocks ξ^r_{jmt}, but such short run adjustments are unlikely. For estimation, we include the baseline voucher and two top-up components related to a school being part of the SEP program and to a school having a concentration of SEP students above a threshold.

We adopt two instruments for school quality. First, we utilize county temperature data on test days. This instrument is motivated by recent research on the relationship between climate and academic achievement (Graff Zivin et al., 2018; Park, 2020). The data supports the existence of a relationship between test-day temperature and test scores, whereas test-day temperature is unlikely to be related to unobserved school attributes.[20] Finally, we use a residualized 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.[21] This instrument would be invalid if the likelihood of winning the prize in a previous year conditional on school quality in that year correlates with current school unobservable attributes.[22]

5.1.3. 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.

[16] 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.

[17] 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 km from the median household location in the market.

[18] Entry and exit of differentiated schools and heterogeneity across schools in a market generate the variation in rival attributes captured by these instruments. In our estimating sample, the yearly school turnover rate has an average and standard deviation of 5% and 12% respectively. This implies that indeed the set of competitors that a school faces in a market changes over time.

[19] We construct school average teacher hourly wages using teacher-level data on earnings and hours worked collected by MINEDUC.

[20] We utilize data from the Berkeley Earth dataset, which provides population-weighted estimates of daily temperature at the county level. We include both temperature and temperature squared to account for non-linear effects of temperature on academic achievement as documented in Graff Zivin et al. (2018). The strength of this instrument decreases when focusing on smaller markets, as the variation in temperature becomes weaker in such cases.

[21] We utilize the residual of a regression of a 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 may be driven by school-specific attributes that could be persistent in time.

[22] Mizala and Urquiola (2013) show that winning a SNED award has no effect on enrollment through an information channel. This result supports the validity of our instrument.

We adopt an approach based on the spatial distance between schools, similar to that in Neilson (2017). Distance has been shown to be a relevant determinant of school choice in the literature (Gallego and Hernando, 2009; Neilson, 2017). In our data, students' average distance to chosen schools is 1.3 miles and the 90th 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, 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.[23]

A description of the resulting sample is displayed in Table 2. The sample is comprised by $R = 2$ household types, $M = 25$ markets, and $T = 4$ years, adding up to 200 cells. The estimating sample includes 1556 schools and 97,471 students. The sample covers an average of 33 percent of the students attending schools in the markets in the sample, and 92 percent of the schools operating in them. Moreover, an average of 49 percent of students included in the sample across markets are eligible for the SEP program.[24]

5.1.4. 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. Our first set of results shows that schools' locations relative to households is an important driver of choices. We estimate that the distance coefficient β_d^r is negative for both low- and high-income households across all markets, reflecting a disutility for choosing a school further away from home.[25] In terms of heterogeneity, low-income households are on average 14 percent more distance-sensitive than high-income households.

Table 3 presents results for different specifications of instrumental variables linear regressions of the estimates of δ_{jmt} on different sets of school characteristics and fixed effects. Columns 1 through 3 display results for all households in the sample, columns 4 through 6 display results for low-income households, and columns 6 through 9 for high-income households. Overall, results point in the expected direction: household utility decreases with school fees and increases with their reported quality. Adding market-year fixed effects and other school attributes to the regression increases the magnitude of point estimates relative to the baseline case.[26] Overall, the model provides a good fit for observed enrollment shares. The correlation between observed and predicted enrollment shares is 0.88.[27]

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. However, estimates for both groups imply sizable responses to school quality, with average quality elasticities of 3.73 and 6.03, respectively. Together, these results imply 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 in the school choice literature (e.g., Gallego and Hernando, 2009; Hastings et al., 2009; Neilson, 2017).[28]

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 δ^r and β_d^r, 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 β_k^r from the second stage of the school choice model, and a counterfactual vector of school characteristics $\widetilde{X}_{ij} = [p_j, \widetilde{q}_j, W_j]$, where $\widetilde{q}_j$ stands for the undistorted quality of school j. For this analysis, we utilize the results for the second stage from our preferred specifications: columns 6 and 9 of Table 3.

5.2.1. School 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 (6) of our school choice model and using parameter estimates and data on school attributes for both scenarios. Following equation (6), choice probabilities are therefore computed

[23] The map presented in Fig. A.7 provides an example for the resulting market definitions, and Table A.2 displays its summary statistics. 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.

[24] 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.

[25] Fig. A.8 displays the estimated coefficients in each market for distance between households and schools for both low- and high-income households.

[26] Tables A.3 and A.4 display results from the first stage of the IV estimation for school fees and quality respectively. The bottom rows in Table 3 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.

[27] For illustration, Fig. A.9 displays observed and simulated enrollment shares for all schools in the sample.

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

Table 2
Summary statistics for estimation of school choice model.

Variable		Mean	SD	p10	p50	p90
Students	In sample	1009	844	324	665	2446
	Coverage rate	0.33	0.12	0.17	0.31	0.48
Schools	In sample	63	62	19	45	134
	Coverage rate	0.92	0.13	0.72	0.97	1.00
Low-income students	In sample	479	391	166	323	1184
	Sample share	0.49	0.11	0.35	0.50	0.60

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

Table 3
IV results from the second stage of school choice model.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All			Low-income students			High-income students		
Fee	−0.003***	−0.004***	−0.006***	−0.006***	−0.007***	−0.010***	−0.002***	−0.003***	−0.005***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)
Quality	0.012***	0.019***	0.019***	0.004**	0.011***	0.015***	0.021***	0.028***	0.023***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Religious			−0.054**			−0.086***			−0.019
			(0.024)			(0.030)			(0.029)
Gender constraint			0.148***			0.121**			0.161***
			(0.047)			(0.059)			(0.055)
Public			0.089***			0.229***			−0.061
			(0.031)			(0.040)			(0.039)
SEP school			−0.325***			−0.533***			−0.587***
			(0.065)			(0.094)			(0.066)
Market-year F.E.	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Observations	10,774	10,774	10,774	5335	5335	5335	5439	5439	5439
First stage tests									
F-test fee	1566.15	2031.73	395.15	484.55	582.07	73.91	1285.81	1593.35	329.15
F-test quality	70.62	17.69	15.03	33.81	9.68	8.17	36.50	8.03	6.86
Cragg-Donald EV	283.97	232.38	203.91	146.05	127.12	101.57	139.36	106.47	98.52

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

as $P^r_{ijmt}(d^r, \hat{\delta}^r, \hat{\beta}^r_d)$ and $P^r_{ijmt}(d^r, \tilde{\delta}^r, \hat{\beta}^r_d)$, where $\tilde{\delta}^r_{jmt} = \sum_k \tilde{x}_{k,jmt} \hat{\beta}^r_k + \hat{\xi}^r_{jmt}$ is the mean utility of school j in market m in period t, computed using preferences estimates and data on counterfactual school quality.

Changes in the quality disclosure system affect household choices. Fig. 4A and B display computed changes in choice probabilities between both scenarios. There is significant heterogeneity, despite the fact that the magnitude of estimated distortions is moderate.[29] 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.[30] 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

[29] This pattern holds when restricting the analysis to the set of schools actually chosen by parents as displayed by Fig. 4C and D.

[30] To put these results in context, we compare them to those in recent studies of information interventions for school choice. Andrabi et al. (2017) find noisy evidence of switching, by which students in villages that receive school report cards switch 5.6% more often to high quality private schools and switch 7.4% less often away from low quality schools. Allende (2012) finds that the ETL information intervention we describe in section 2.2 induced local net increases in enrollment around the cutoffs used for school classification of about 1.5%. Finally, and less related to our context and outcome variable, Hastings and Weinstein (2008) find that informed households apply to a non-guaranteed school 22.5% more than uninformed households. Overall, the changes in school choice behavior that we estimate are in the range of these previous estimates.

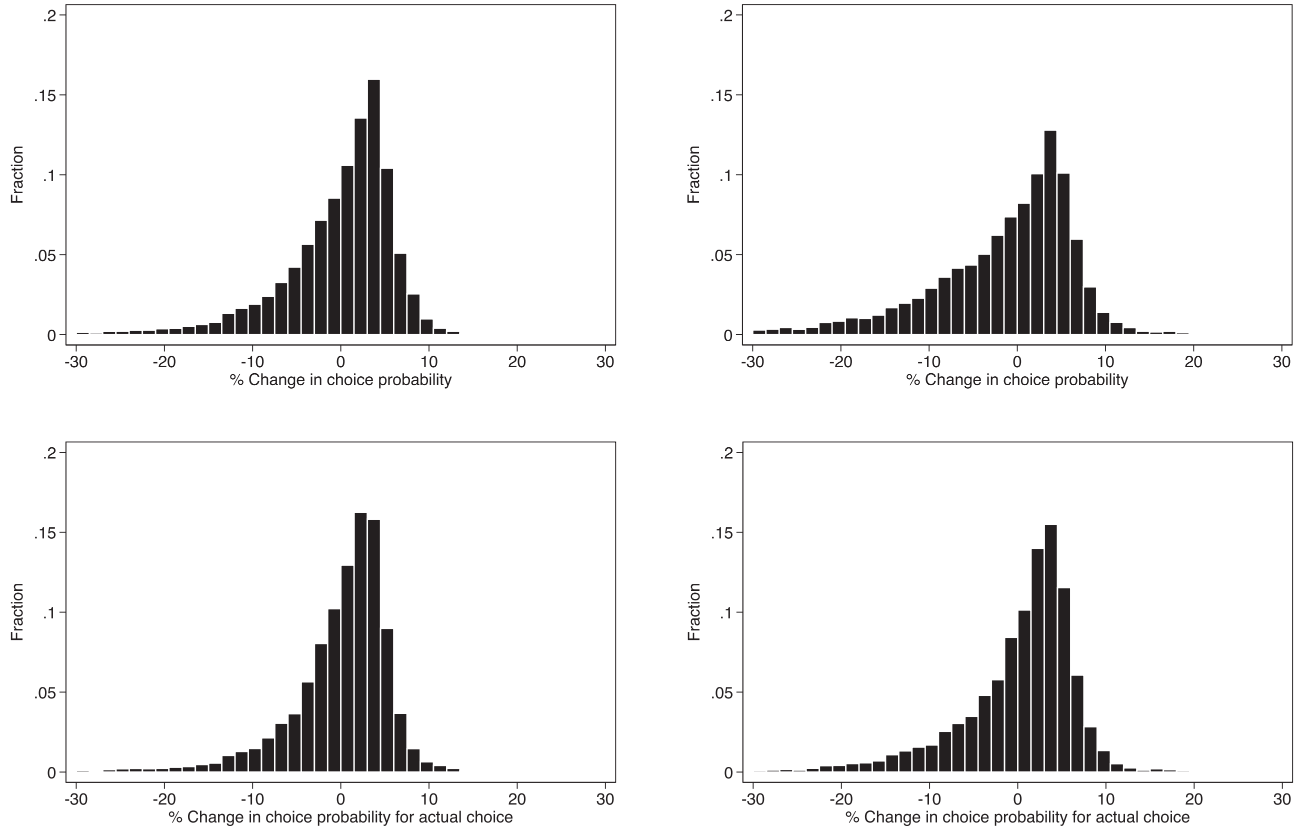

Fig. 4. Changes in choice probabilities. (a) Low-income students, all schools; (b) High-income students, all schools; (c) Low-income students, chosen school; (d) High-income students, chosen school.
Notes: These figures display change in school choice probabilities between the counterfactual and baseline scenarios we analyze. Each observation is the percentage change in the choice probability of a school by a household in the estimating dataset. Panels (a) and (b) include results for all schools in the dataset, while panels (c) and (d) focus only on schools chosen by household in the baseline scenario.

switchers.[31]

We compute the predicted attributes of schools chosen by households under both scenarios. Table 4 displays results for low- and high-income households. We report the average across switchers and across all households. Columns 1 and 3 in Table 4 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 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 switchers would change their choices substantially. 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 4 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.

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 choice and experience utility coincide.

Let u_{ij} be the utility of household i from school j under distorted school quality, choice utility. Similarly, let $\widetilde{u}_{ij}$ 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 $\widetilde{u}_{ij} = u_{ij} + \tau_j$, where τ_j measures the wedge between choice and experienced utility from school j. Under the utility function assumed in section 5.1, we know that $\tau_j = \beta_q(\widetilde{q}_j - q_j)$.[32]

[31] We 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.

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

Table 4
Means of predicted school attributes of households choices.

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

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

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

$$j_i^* = \arg\max_j \{u_{ij}\}_{j \in \mathcal{J}_i}$$

$$\widetilde{j}_i^* = \arg\max_j \{\widetilde{u}_{ij}\}_{j \in \mathcal{J}_i}$$

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

$$E[\Delta CS_i] = \frac{1}{\beta_p}\left[\log\sum_j \exp(\widetilde{v}_{ij}) - \log\sum_j \exp(v_{ij}) - \sum_j P_{ij}\tau_j\right] \tag{8}$$

where we define $\widetilde{v}_{ij} \equiv \widetilde{\delta}_j + \beta_d d_{ij}$ 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 β_p simply transforms the welfare loss to monetary units. Equation (8) calculates the average gain in consumer surplus across all households in the sample. We can then compute average gains in consumer surplus for switchers or aggregate these gains across different dimensions. These welfare gains can alternatively be interpreted as the average willingness to pay of households for undistorted quality information.

Results from welfare calculations are displayed by Table 5 and show that expected welfare would increase in the counterfactual scenario for all households. 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.[33]

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 5. 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.[34]

Second, we let low-income households have the same preferences as high-income households on all school attributes. The share of switchers increases by 0.6 percentage point to 3.8 percent. Average yearly gains for low-income switchers in this counterfactual would climb to \$181, more than three times those in the first counterfactual and higher than those for high-income switchers.[35] These results imply that differences in preferences are enough to explain the gap across groups in welfare

[33] 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.

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

[35] The fact that welfare gains for the low-income when endowed with preferences of high-income households are larger than those when endowed with such preference only over school quality comes partly from the fact that we estimate high-income households to be less price-sensitive. This implies that the willingness to pay for a given increase in quality is higher than under low-income preferences as can be noted in equation (8).

Table 5
Yearly welfare gains of providing undistorted quality signals.

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

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

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 information on undistorted quality signals is provided to households. Results point in three directions. First, distortions in quality signals have effects on choices, as choice probabilities 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 beneficial to invest in reducing distortions of quality disclosure systems in educational markets. Note that our counterfactual exercise does not evaluate the effects of the overall quality disclosure system or of a particular information intervention to households, but rather the effects of correcting a particular distortion to school quality signals given the disclosure system in place. 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.

Finally, there are some limitations to this counterfactual analysis that we acknowledge. Our analysis relaxes the assumption of full information by households which has been commonly imposed in the school choice literature. However, we focus on a particular information friction while assuming that households are informed about most other dimensions, including the set of schools available in the market, their attributes and other components of their quality than the distortions we focus on. Therefore, we interpret our analysis as measuring the effect of improving an already sophisticated quality disclosure system. Moreover, we assume away any other frictions or biases that households may face when choosing schools. Relaxing these assumptions is a relevant line of work for better understanding the role of quality disclosure policies on school choice behavior more broadly. Recent work by Allende et al. (2019) and Kapor et al. (2020) makes progress on this direction by developing richer models of school choice under imperfect information.

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

In 2012 the total amount of public resources transferred to schools in the form of teacher bonuses through the SNED program reached 15 million U.S. dollars. The sharp discontinuity to assigning resources is based on the following index for each school:

$$I_{jgt}(q_{j\tau}, q_{j\tau-1}, X_{j\tau}) = \omega_1 q_{j\tau} + \omega_2(q_{j\tau} - q_{j\tau-1}) + \boldsymbol{\omega}_3' \mathbf{X}_{j\tau} \tag{9}$$

where I_{jgt} is the index of school j, in group g, and year t; $q_{j\tau}$ is the average test score in year τ; $\mathbf{X}_{j\tau}$ is a vector of attributes; and $(\omega_1, \omega_2, \boldsymbol{\omega}_3)$ are weights chosen by the government, with $\omega_k \in (0,1)$ and $\sum_k \omega_k = 1$. More precisely, $\omega_1 = 0.37$ and $\omega_2 = 0.28$, thus test scores weight 65% in the formula for the index. 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 $(\widetilde{q}_{j\tau}, \widetilde{q}_{j\tau-1})$, we calculate

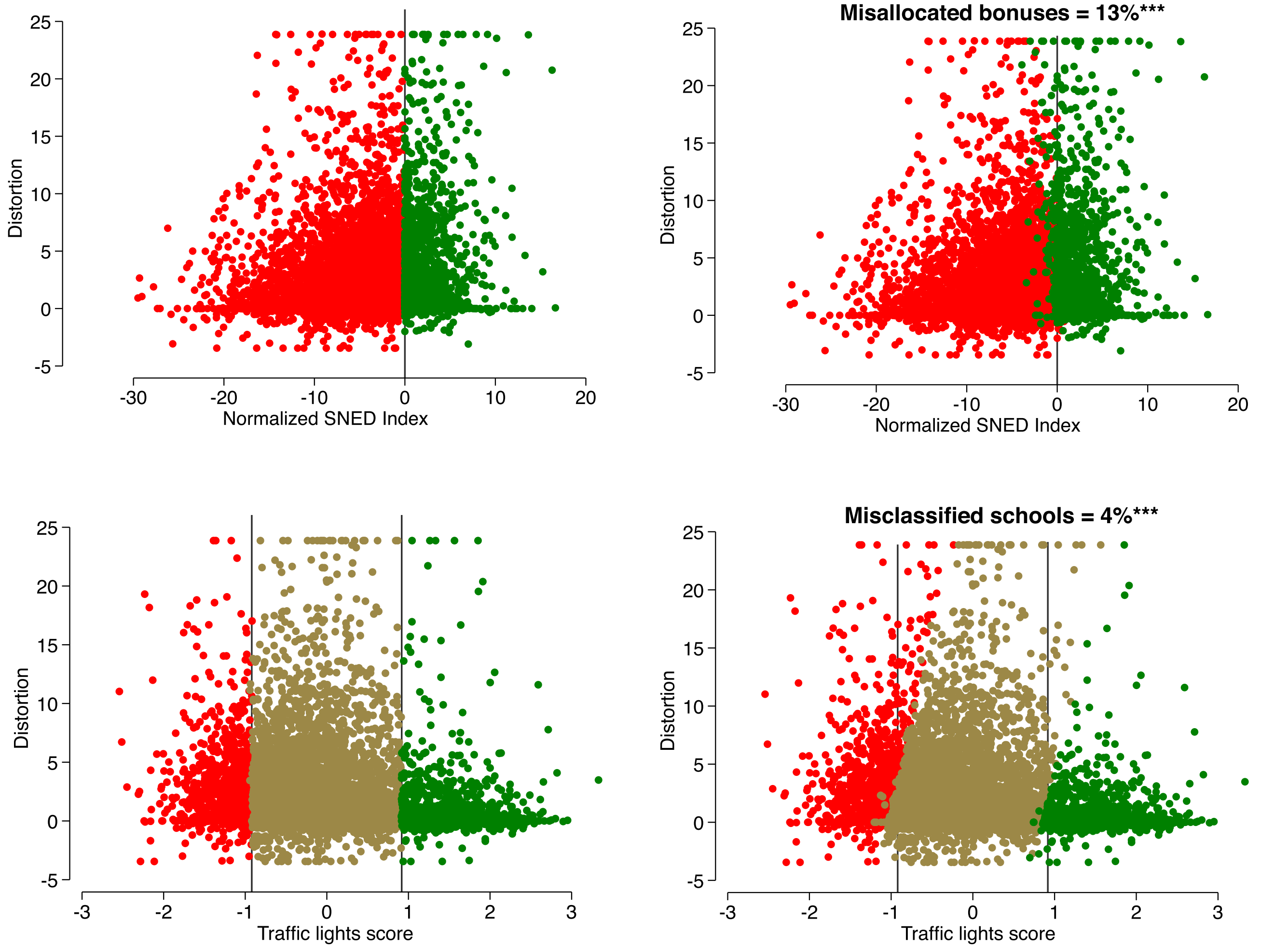

Fig. 5. Misallocation of public programs. (a) Teacher bonuses (actual assignment). (b) Teacher bonuses (counterfactual). (c) Information (actual provision). (d) Information (counterfactual).

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

schools undistorted indices using equation (9), $\widetilde{I}_{jgt} = I_{jgt}(\widetilde{q}_{j\tau}, \widetilde{q}_{j\tau-1}, \mathbf{X}_{j\tau})$, and reallocate bonuses based on these undistorted measures.

Fig. 5a and 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 1000 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 1000 times. Bounds for our misallocation estimates can be constructed using the estimated distribution of misallocation.

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

6.2. Information for school choice

In 2010 schools were classified into three mutually exclusive categories as part of the “Educational Traffic Lights” program. 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:

$$c_j(q_{jt}) = r \cdot 1[q_{jt} < \underline{s}] + y \cdot 1[\underline{s} < q_{jt} < \overline{s}] + g \cdot 1[q_{jt} > \overline{s}] \quad (10)$$

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

Equation (10) 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 $\widetilde{c}_j = c_j(\widetilde{q}_{jt})$, 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.

Fig. 5c and 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 5000 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, where high-stakes testing is a relevant component of the accountability system. In particular, we have quantified how non-random attendance on test day causes school quality signals to be distorted. We study the determinants of these distortions and we find that they are largely explained by fixed school characteristics and that strategic behavioral responses by schools play a role. The latter is consistent with the so-called Campbell's Law: the higher the stakes are for an indicator of a social phenomenon, the more liable it is to be distorted (Campbell, 1979). Distortions, however, are not *per se* a reason of concern. To claim distortions have costs, we need to study the impacts they have on decisions. The Chilean market-oriented educational system is particularly interesting to study such impacts because test scores are not just used for the two objectives of quality assessment and performance evaluation emphasized by Neal (2013), but rather for three, as they also feed school choice. As we have shown that distortions have negative impacts on school choice and induce misallocation of public programs, we conclude that distortions can impose significant costs in educational markets.

Our study is, to the best of our knowledge, the first to quantify the *market* consequences from distortions in quality signals. Further research is required to quantify other distortions and to address other margins of educational markets. We highlight that the institutional environment might determine the magnitude and impacts of distortions. 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 for developing countries constructing accountability systems, and also for developed countries with imperfect systems already in place. 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 NCLB). We also highlight that imputing low test scores as a penalty could harm schools serving disadvantaged populations that have low attendance rates for reasons unrelated to schools. In addition, we emphasize that the magnitude of elasticities determines the extent to which households can benefit from information policies. In school markets, we argue that complementary policies that increase quality-sensitivity of low-income households might enable them to benefit more from accurate information. Finally, our results on misallocation of public programs provide an argument against sharp assignment rules for public programs based on variables prone to distortions. Multiple programs in different countries and sectors are assigned through such rules and might be subject to misallocation.

Authors statement

All co-authors of this paper declare to have no relevant or material financial interests that relate to the research described in this paper.

Appendix A. Additional results

Additional results to this article can be found online at https://doi.org/10.1016/j.jdeveco.2020.102532.

References

Alfaro, R., Fuenzalida, M., 2009. Imputación Múltiple en Encuestas Microeconómicas. Cuad. Econ. 46, 273–288.

Allende, C., 2012. The Impact of Information on Academic Achievement and School Choice: Evidence from Chilean "Traffic Lights". Manuscript. .

Allende, C., Gallego, F., Neilson, C., 2019. Approximating the Equilibrium Effects of Informed School Choice. Princeton University Industrial Relations Section. Working Paper 628.

Andrabi, T., Das, V., Khwaja, A., 2017. Report cards: the impact of providing school and child test scores on education markets. Am. Econ. Rev. 107 (6), 1535–1563.

Apperson, J., Bueno, C., Sass, T.R., 2016. Do the Cheated Ever Prosper? the Long-Run Effects of Test-Score Manipulation by Teachers on Student Outcomes.

ASER, 2019. ASER 2018-Rural. ASER Centre.

Bagwell, K., 2007. The economic analysis of advertising. In: Armstrong, M., Porter, R. (Eds.), Handbook of Industrial Organization, vol. 3. Elsevier, pp. 1701–1844.

Barrera-Osorio, F., Bertrand, M., Linden, L.L., Perez-Calle, F., 2011. Improving the design of conditional transfer programs: evidence from a randomized education experiment in Colombia. Am. Econ. J. Appl. Econ. 3 (2), 167–195.

Bau, N., 2019. Estimating an Equilibrium Model of Horizontal Competition in Education. Manuscript.

Baum, D., Lewis, L., Lusk-Stover, O., Patrinos, H., 2014. What Matters Most for Engaging the Private Sector in Education? World Bank Framework Paper. .

Bayer, P., Ferreira, F., McMillan, R., 2007. A unified framework for measuring preferences for schools and neighborhoods. J. Polit. Econ. 115 (4), 588–638.

Bellei, C., 2009. Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile. Econ. Educ. Rev. 28 (5), 629–640.

Bernheim, B.D., Rangel, A., 2009. Beyond revealed preference: choice-theoretic foundations for behavioral welfare Economics. Q. J. Econ. 124 (1), 51.

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

Campbell, D., 1979. Assessing the impact of planned social change. Eval. Progr. Plann. 2, 67–90.

Carneiro, P., Das, J., Reis, H., 2019. The Value of Private Schools: Evidence from Pakistan.

Centro de Investigación y Desarrollo de la Educación, 2010. VIII Encuesta a Actores del Sistema Educativo 2010.

Chay, K.Y., McEwan, P.J., Urquiola, M., 2005. The central role of noise in evaluating interventions that use test scores to rank schools. Am. Econ. Rev. 95 (4), 1237–1258.

Chumacero, R.A., Gómez, D., Paredes, R., 2011. I would walk 500 miles (if it paid): vouchers and school choice in Chile. Econ. Educ. Rev. 30, 1103–1114.

Contreras, D., Rau, T., 2012. Tournament incentives for teachers: evidence from a scaled-up intervention in Chile. Econ. Dev. Cult. Change 61 (1), 219–246.

Correa, J.A., Parro, F., Reyes, L., 2014. The effects of vouchers on school results: evidence from Chile's targeted voucher program. J. Hum. Cap. 8 (4), 351–398.

Cragg, J.G., Donald, S.G., 1993. Testing identifiability and specification in instrumental variable models. Econom. Theor. 9, 222–240.

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

Dee, T.S., Dobbie, W., Jacob, B.A., Rockoff, J., 2019. The causes and consequences of test score manipulation: evidence from the New York regents examinations. Am. Econ. J. Appl. Econ. 11 (3), 382–423.

Deming, D.J., Cohodes, S., Jennings, J., Jencks, C., 2016. School accountability, postsecondary attainment and earnings. Rev. Econ. Stat. 98 (5), 848–862.

Diamond, R., Persson, P., 2017. The Long-Term Consequences of Teacher Discretion in Grading of High-Stakes Tests. Manuscript. .

Dorfman, R., Steiner, P.O., 1954. Optimal advertising and optimal quality. Am. Econ. Rev. 44 (5), 826–836.

Dranove, D., Jin, G.Z., 2010. Quality disclosure and certification: theory and practice. J. Econ. Lit. 48 (4), 935–963.

Dranove, D., Kessler, D., McClellan, M., Satterthwaite, M., 2003. Is more information better? The effects of report cards on health care providers. J. Polit. Econ. 111 (3), 555–588.

Ebenstein, A., Lavy, V., Roth, S., 2016. The long run consequences of high-stakes examinations: evidence from transitory variation in pollution. Am. Econ. J. Appl. Econ. 8 (4), 36–65.

Feigenberg, B., Rivkin, S., Yan, R., 2018. Illusory Gains from Chile's Targeted School Voucher Experiment. National Bureau of Economic Research. Working Paper 23178.

Figlio, D., Loeb, S., 2011. School accountability. In: Eric, A., Hanushek, S.M., Woessmann, L. (Eds.), Handbook of the Economics of Education, vol. 3. Elsevier, pp. 383–421.

Figlio, D., Winicki, J., 2005. Food for thought: the effects of school accountability plans on school nutrition. J. Publ. Econ. 89 (2–3), 381–394.

Figlio, D.N., 2006. Testing, crime and punishment. J. Publ. Econ. 90 (4–5), 837–851.

Figlio, D.N., Getzler, L.S., 2002. Accountability, Ability and Disability: Gaming the System. National Bureau of Economic Research. Working Paper 9307.

Fryer, R., 2017. Chapter 2 - the production of human capital in developed countries: evidence from 196 randomized field experiments. In: Banerjee, A., Duflo, E. (Eds.), Handbook of Economic Field Experiments, Volume 2 of Handbook of Economic Field Experiments, , pp. 95–322 North-Holland.
Gallego, F., Hernando, A., 2009. School Choice in Chile: Looking at the Demand Side. Pontificia Universidad Catolica de Chile Documento de Trabajo (356).
Glewwe, P., Muralidharan, K., 2016. Chapter 10 - Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications. Volume 5 of Handbook of the Economics of Education. Elsevier, pp. 653–743.
Gómez, D., Chumacero, R., Paredes, R., 2012. School choice and information. Estud. Econ. 39 (2), 143–157.
Graff Zivin, J., Hsiang, S., Neidell, M., 2018. Temperature and human capital in the short and long run. J. Assoc. Environ. Resour. Econ. 5 (1), 77–105.
Hastings, J., Kane, T.J., Staiger, D.O., 2009. Heterogeneous Preferences and the Efficacy of Public School Choice. Manuscript.
Hastings, J.S., Weinstein, J.M., 2008. Information, school choice, and academic achievement: evidence from two experiments. Q. J. Econ. 123 (4), 1373–1414.
Jacob, B.A., 2005. Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago public schools. J. Publ. Econ. 89, 761–796.
Jacob, B.A., Levitt, S.D., 2003. Rotten apples: an investigation of the prevalence and predictors of teacher cheating. Q. J. Econ. 118 (3), 843–877.
Jensen, R., 2010. The (perceived) returns to education and the demand for schooling. Q. J. Econ. 125 (2), 515–548.
Jin, G.Z., Leslie, P., 2003. The effect of information on product quality: evidence from restaurant hygiene grade cards. Q. J. Econ. 118 (2), 409–451.
Kane, T.J., Staiger, D.O., 2002. The promise and pitfalls of using imprecise school accountability measures. J. Econ. Perspect. 16 (4), 91–114.
Kapor, A.J., Neilson, C.A., Zimmerman, S.D., 2020. Heterogeneous beliefs and school choice mechanisms. Am. Econ. Rev. 110 (5), 1274–1315.
Kennickell, A.B., 1998. Multiple Imputation in the Survey of Consumer Finances. Woking Paper. .
Kremer, M., Holla, A., 2009. Improving education in the developing world: what have we learned from randomized evaluations? Ann. Rev. Econ. 1 (1), 513–542.
Lemke, R.J., Hoerandner, C.M., McMahon, R.E., 2006. Student assessments, non-test-takers, and school accountability. Educ. Econ. 14 (2), 235–250.
Mas, A., Moretti, E., 2009. Peers at work. Am. Econ. Rev. 99 (1), 112–145.
Mbiti, I.M., 2016. The need for accountability in education in developing countries. J. Econ. Perspect. 30 (3), 109–132.
McEwan, P.J., Urquiola, M., Vegas, E., 2008. School Choice, Stratification and Information on School Performance: Lessons from Chile. Economia.
Meghir, C., Rivkin, S., 2011. Econometric Methods for Research in Education. Volume 3 of Handbook of the Economics of Education. Elsevier, pp. 1–87.
Ministry of Education, 2013. Estadísticas de la Educación 2013.
Mizala, A., Urquiola, M., 2013. School markets: the impact of information approximating schools' effectiveness. J. Dev. Econ. 103, 313–335 0.
Muralidharan, K., 2017. Chapter 3 - field experiments in education in developing countries. In: Banerjee, A., Duflo, E. (Eds.), Handbook of Economic Field Experiments, Volume 2 of Handbook of Economic Field Experiments, , pp. 323–385 North-Holland.
Muralidharan, K., Sundararaman, V., 2015. The aggregate effect of school choice: evidence from a two-stage experiment in India. Q. J. Econ. 130 (3), 1011–1066.
NCES, 2011. vert. https://nces.ed.gov/programs/digest/d11/tables/dt11_043.aspvert.
Neal, D., 2013. The consequences of using one assessment system to pursue two objectives. J. Econ. Educ. 44 (4), 339–352.
Neal, D., Schanzenbach, D.W., 2010. Left behind by design: proficiency counts and test-based accountability. Rev. Econ. Stat. 92 (2), 263–283.
Neilson, C., 2017. Targeted Vouchers, Competition Among Schools, and the. Academic Achievement of Poor Students, Manuscript.
Nelson, P., 1970. Information and consumer behavior. J. Polit. Econ. 78 (2), 311–329.
Park, J., 2020. Hot temperature and high stakes performance. J. Hum. Resour. (Forthcoming).
Quezada-Hofflinger, A., Von Hippel, P.T., 2018. The Response to High Stakes Testing in Chile, 2005-2013: Legitimate and Illegitimate Ways to Raise Test Scores. Manuscript. .
Rockoff, J., Turner, L.J., 2010. Short-run impacts of accountability on school quality. Am. Econ. J. Econ. Pol. 2 (4), 119–147.
Rubin, D.B., 1987. Multiple Imputation for Nonresponse in Surveys. Wiley & Sons, New York.
Sánchez, C., 2018. Targeted or Universal? Mobilizing Students through School Vouchers. Manuscript. .
Sánchez, C., 2019. Skipping Your Exam? the Unexpected Response to a Targeted Voucher Policy. Manuscript. .
Small, K., Rosen, H., 1981. Applied welfare Economics with discrete choice models. Econometrica 49 (1), 105–130.
Stock, J., Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, pp. 80–108 chapter 5.
Train, K., 2015. Welfare calculations in discrete choice models when anticipated and experienced attributes differ: a guide with examples. J. Choice Modell. 16, 15–22.
Walters, C., 2018. The demand for effective charter schools. J. Polit. Econ. 126 (6), 2179–2223.