No.23

JULIO DE 2019

Documentos **CEDE**

ISSN 1657-7191 Edición electrónica.

Telling schools apart: the role of preferences,
constraints, and the ability to differentiate
between schools in parents' choices

Diego Amador
Juan-Andrés Castro
Nicolás Grau

CEDE
CENTRO DE ESTUDIOS SOBRE DESARROLLO ECONÓMICO

**Universidad de
los Andes**
Facultad de Economía

Serie Documentos Cede, 2019-23
ISSN 1657-7191 Edición electrónica.
Julio de 2019

© 2019, Universidad de los Andes, Facultad de Economía,
CEDE. Calle 19A No. 1 – 37 Este, Bloque W.
Bogotá, D. C., Colombia Teléfonos: 3394949- 3394999,
extensiones 2400, 2049, 2467
infocede@uniandes.edu.co
<http://economia.uniandes.edu.co>

Impreso en Colombia – Printed in Colombia

La serie de Documentos de Trabajo CEDE se circula con propósitos de discusión y divulgación. Los artículos no han sido evaluados por pares ni sujetos a ningún tipo de evaluación formal por parte del equipo de trabajo del CEDE.

El contenido de la presente publicación se encuentra protegido por las normas internacionales y nacionales vigentes sobre propiedad intelectual, por tanto su utilización, reproducción, comunicación pública, transformación, distribución, alquiler, préstamo público e importación, total o parcial, en todo o en parte, en formato impreso, digital o en cualquier formato conocido o por conocer, se encuentran prohibidos, y sólo serán lícitos en la medida en que se cuente con la autorización previa y expresa por escrito del autor o titular. Las limitaciones y excepciones al Derecho de Autor, sólo serán aplicables en la medida en que se den dentro de los denominados Usos Honrados (Fair use), estén previa y expresamente establecidas, no causen un grave e injustificado perjuicio a los intereses legítimos del autor o titular, y no atenten contra la normal explotación de la obra.

Universidad de los Andes | Vigilada Mineducación
Reconocimiento como Universidad: Decreto 1297 del 30 de mayo de 1964. Reconocimiento personería jurídica: Resolución 28 del 23 de febrero de 1949 Minjusticia.

Telling schools apart: the role of preferences, constraints, and the ability to differentiate between schools in parents' choices^{*}

Diego Amador^{**}

Juan-Andrés Castro^{***}

Nicolás Grau^{****}

Abstract

Limitations in the ability of parents to compare schools have important implications for market oriented educational systems, which rely on parents choices to improve quality through competition. To empirically study these limitations, we develop and estimate a static model of elementary school choice that distinguishes between preferences for academic quality, the ability to differentiate between schools of different quality, and constraints in terms of the schools available to different households. Because school quality might be endogenous to parents' choices, we identify the key parameters related to preferences for quality using exogenous variation in schools' funding introduced by a policy that substantially increased the voucher amount for each enrolled socioeconomically vulnerable student. We estimate the model using a combination of administrative and survey data from Chile, which includes rich information on how parents compare the academic quality of schools. Using counterfactual simulations, we find the interaction between limitations to tell schools apart and differences in preferences across households plays an important role in decreasing the quality of schools attended by Chilean children, especially for children with less educated parents.

Keywords. School choice, structural estimation, bounded rationality, education systems. **JEL Classification:** I24, C35.

^{*}We thank the Chilean Ministry of Education for providing the data. We are grateful to Naoki Aizawa, Raquel Bernal, Chao Fu, Holger Sieg and to seminar participants at the Workshop on Applied Economics of Education (Catanzaro, 2019), LACEA-LAMES 2018, Center for Applied Economics (Universidad de Chile), Universidad del Rosario, and CEDE (Universidad de los Andes) for helpful comments and suggestions. We thank the Vicerrectoría de Investigaciones - CEDE fund at Universidad de los Andes for financial support. Nicolás Grau thanks the Centre for Social Conflict and Cohesion Studies (CONICYT-FONDAP-15130009) for financial support. Powered@NLHPC: This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02). All remaining errors are our own.

^{**}Texas Policy Lab, Rice University, and Department of Economics, Universidad de los Andes.

^{***}Central Bank of Chile.

^{****}Department of Economics, Faculty of Economics and Business, University of Chile. Corresponding author (nicolas.grau@gmail.com).

Comparando escuelas: el rol de las preferencias, restricciones y las habilidades para diferenciar escuelas en las decisiones de los padres.^{*}

Diego Amador^{**}

Juan-Andrés Castro^{***}

Nicolás Grau^{****}

Resumen

Las limitaciones en la habilidad de los hogares para comparar escuelas tienen implicaciones importantes para sistemas educativos con una orientación de mercado, los cuales se sostienen en las decisiones de los hogares para mejorar la calidad académica por medio de la competencia. Para estudiar estas limitaciones desde una perspectiva empírica, desarrollamos y estimamos un modelo estático de decisiones de educación primaria que distingue entre preferencias por calidad académica, restricciones en la capacidad de diferenciar las escuelas en términos de dicha calidad y restricciones en cuanto al conjunto de colegios disponibles para cada hogar. Debido a que la calidad académica puede ser endógena a las decisiones de los hogares, hacemos uso de la variación exógena creada por una política que alteró la financiación de las escuelas con base en la matrícula de estudiantes de condiciones vulnerables para identificar los parámetros del modelo. Estimamos el modelo usando una combinación de datos administrativos y encuestas en Chile, los cuales incluyen información detallada acerca de la forma en que los padres comparan la calidad académica de las escuelas. Por medio de simulaciones contrafactuales, encontramos que la interacción entre las limitaciones para diferenciar escuelas y la heterogeneidad en preferencias entre los hogares juegan un papel importante en reducir la calidad académica a la cual acceden los niños chilenos, especialmente para los hijos de padres menos educados.

Palabras clave. Elección escolar, estimación estructural, racionalidad limitada, sistemas educativos. **Clasificació JEL :** I24, C35.

^{*} Agradecemos al Ministerio de Educación de Chile por permitirnos hacer uso de los datos para este artículo. Estamos muy agradecidos con Naoki Aizawa, Raquel Bernal, Chao Fu, Holger Sieg y participantes del Workshop on Applied Economics of Education (Catanzaro, 2019), LACEA-LAMES 2018, y seminarios en el Centro de Economía Aplicada (Universidad de Chile), Universidad del Rosario, y CEDE (Universidad de los Andes) por sus comentarios y sugerencias. Agradecemos también a la convocatoria Vicerrectoría de Investigaciones - CEDE en la Universidad de los Andes por el apoyo financiero. Nicolás Grau agradece al Centro de Estudios para el Conflicto y la Cohesión Social(CONICYT-FONDAP-15130009) por el apoyo financiero. Powered@NLHPC: Esta investigación estuvo apoyada por la infraestructura de supercomputadoras NLHPC (ECM-02). Todos los errores son exclusivamente de los autores.

^{**} Texas Policy Lab, Rice University, y Departamento de Economía, Universidad de los Andes.

^{***} Banco Central de Chile.

^{****} Departamento de Economía, Facultad de Economía y Negocios, Universidad de Chile. Contacto (nicolas.grau@gmail.com).

1. Introduction

A frequent topic in policy debates is what should be the role – if any – of market incentives in education provision. Because parents’ choices are the critical mechanism to increase school quality in a market-oriented educational system, the literature has focused on the extent to which parents consider school quality when they make their decisions, and how this consideration is heterogeneous across parents. However, to understand parents’ school choices, and the potential heterogeneity in their preferences, one must separate the effects of differences in their preferences, in their ability to distinguish schools in terms of their quality, and in their choice sets. Distinguishing these three elements is a complex empirical task, as, normally, one would only observe which school is attended by any given student, a non trivial outcome of the interaction of these three elements. Moreover, the fact that school quality is potentially endogenous to parents’ school choices further complicates this task.

In this paper, we build a simple static model of elementary school choice that explicitly distinguishes between differences in preferences for quality, in perceptions about quality, and in choice sets. To estimate the model, we combine detailed Chilean administrative data with newly available survey data from the Encuesta de Elección Escolar (EEE), which focuses on how parents compare and choose schools. We use the rich information from the EEE to build a measure of parents’ ability to compare the relative quality of different schools, as measured by results in standardized achievement tests, which are public and widely advertised in Chile. This measure proves to be crucial in distinguishing between preferences for quality and the ability to differentiate schools in terms of that quality.

The model allows for parental heterogeneity along three dimensions. First, besides other characteristics – such as the school’s socioeconomic composition, religious affiliation, type of administration (private, private with voucher, public), and tuition fee– parents in the model care about academic quality.¹ However, there is heterogeneity in terms of how much they care about this quality. Second, parents also differ in their ability to understand publicly available information about academic quality. Specifically, parents in our model can be more or less able at perceiving the differences in quality between schools available to them. Finally, parents face different choice sets, which might be due to a number of reasons, such as where

¹For a general discussion about the control function approach, see [Navarro \(2010\)](#).

they live or the fact that schools may select students based on academic and non-academic characteristics.² Thus, we use our data to define these different choice sets and include an explicit model of the schools' selection process.

By estimating the structural parameters of the model, we are able to assess the empirical relevance of each of these components in determining the quality of schools attended by Chilean children and how this might differ for families from different socioeconomic backgrounds. As many authors have emphasized (e.g., [Gallego and Hernando \(2008\)](#) and [Hsieh and Urquiola \(2006\)](#)), the Chilean system is probably the most massive school choice program in the world, making it an ideal context to work in to assess the relative importance of these different determinants of school choice. Thus, our results speak to the issue of the potential limitations of market-based school systems in general.

In a voucher system, attracting more students leads to additional funding that in turn could allow schools to increase academic quality. Thus, we worry about academic quality being endogenous in our estimations. To address this issue, we use the exogenous variation in school funding introduced in 2008 by the Ley de Subvención Escolar Preferencial (SEP). The SEP established a new targeted voucher that would complement the existing universal voucher by transferring significantly more resources to schools based on the number of enrolled students coming from socioeconomically vulnerable backgrounds. Given this policy variation, we build an instrument for school quality by calculating the additional amount of funding each school would have received through the SEP, had their socio-economic composition stayed as it was before our sample period (and before the SEP was implemented). This instrument belongs to a family of instruments commonly known as shift-share or Bartik instruments, an approach used for the first time in [Bartik \(1991\)](#) and popularized in [Blanchard and Katz \(1992\)](#).

We estimate the parameters of the model by simulated maximum likelihood, implementing a control function approach to address the potential endogeneity of academic quality. Our data is composed by a sample of students who entered the 1st grade between 2008 and 2015. As mentioned, we use data from several sources, which include the Encuesta de Elección Escolar (EEE), administrative records on the Chilean standardized achievement test (the SIMCE) in 4th grade, information from the parent surveys associated with the SIMCE test, and data

²In Chile, at least in our sample period, schools were allowed to select students based on these characteristics (e.g., parents' marital status, knowledge, religious affiliation).

from the Ministry of Education on schools' religious affiliation and total enrollment. The SIMCE surveys include questions about school tuition fees, as well as information about the criteria considered by the school in the admission process. The EEE, as we have highlighted, provides us with key information for the estimation, based on detailed and thorough questions aimed at measuring how well can parents differentiate between schools of different academic qualities.

The model fits the data well. The estimation results show there is important heterogeneity, both in terms of preferences for quality and in the ability to differentiate between schools of different academic qualities, a result that is confirmed by reduced form estimations. We conduct a series of counterfactual experiments to explore the relative relevance of each of these elements in determining the observed quality for households of different socioeconomic backgrounds. First, we find heterogeneity in preferences and perceptions to play a relatively small role on their own in limiting access to academic quality. However, the interaction of both of these elements proves to be of importance. If all households cared about quality as much as the households who care the most, and all of them had the same ability to differentiate qualities as much as the more able households, average quality would increase by 0.096 standard deviations. These results, as expected, are of higher magnitude for households with lower educated parents, which appear to face the strongest constraints. Our results also show that, while school selectivity does not seem to substantially limit access to academic quality, there is strong evidence to suggest a prominent role of socioeconomic segregation in constraining access to quality for less educated households.

The paper has three main contributions. First, the structural approach we follow in this paper allows us to quantify the relative importance different causes – namely preferences, limitations in information, and restrictions in choice sets – of unequal access to high quality schools. To this end, the combination of a structural approach with the exogenous variation for school quality that stems from our instruments is critical. Second, the model considers the difficulties that parents may have in processing and understanding information about school quality, which contributes to the scarce literature on structural estimation with bounded rationality. Finally, we make use of newly available survey data which, when combined with several sources of administrative data, allows us to credibly separate preferences from

information in the estimation of the model.

The structure of the paper is as follows: Section 2 presents a review of the related literature; Section 3 briefly describes the Chilean educational system; Section 4 describes the sources of data and presents descriptive statistics of the estimation sample; Section 5 presents the results of the reduced form estimations; Section 6 introduces the model for the structural estimation; Section 7 describes the estimation method; Section 8 presents the estimated results and the model fit; and Section 9 presents the counterfactual experiments. Section 10 concludes.

2. Literature Review

This paper is related to several strands in the literature. First of all, it is related to the papers that evaluate the role of competitive market incentives in education provision. On one side, there are theoretical papers, such as [Epple and Romano \(1998\)](#), and [McMillan \(2004\)](#), which debate the potential for those incentives, specifically tuition vouchers, to increase schools' quality and to make significant improvements for poor families.³ On the other side, there are empirical studies that show mixed evidence regarding this debate; see, for example, [Hsieh and Urquiola \(2006\)](#), and [Angrist et al. \(2006\)](#).⁴ Overall, as [Epple et al. \(2017\)](#) emphasize, this literature is characterized by mixed evidence, which explain why they argue that the evidence to date is not sufficient to warrant recommending that vouchers (i.e. school competition) be adopted on a widespread basis. There are several studies that try to study the effect of the voucher system implementation in Chile. Although the evidence is mixed regarding its effect on school quality, there is more agreement on the negative effect of this policy on student socioeconomic segregation ([Auguste and Valenzuela \(2006\)](#); [Gauri \(1999\)](#); and [Hsieh and Urquiola \(2006\)](#)). Considering a broader definition of school quality, [Bravo et al. \(2010\)](#) find that educational vouchers increased educational attainment, high school graduation, college attendance and graduation, and wages; whereas [Gajardo and Grau \(2019\)](#) find that even though competition among schools generates a moderate increase in

³In a survey of this literature, [Epple and Romano \(2012\)](#) conclude: *Research taking account of distinctive features of the education "market" has shown that early arguments touting the virtues of laissez-faire flat-rate vouchers were overly optimistic. However, the research does not vindicate voucher opponents who use shortcomings of the laissez-faire voucher to justify the wholesale dismissal of vouchers.*

⁴[Bettinger \(2011\)](#) reviews the cases of Chile, Colombia, and Sweden, emphasizing the context-specific nature of the results.

standardized test results, it has a negative impacts on academic self-esteem and motivation, the climate within the school, civic participation, and healthy lifestyle habits.

Many authors have studied the determinants of parents' school choice because this is one of the most important mechanisms that could explain the shortcomings in the implementation of market-oriented policies in education. In particular, they have studied whether and to what extent parents consider school quality when they choose among schools.⁵ For instance, [Hastings and Weinstein \(2008\)](#) used combined evidence from a natural experiment and a field experiment that provided direct information on school test scores to lower-income families in a public school choice plan, finding a significant increase in the probability that those families would choose higher-performing schools. In an alternative strategy, several studies have focused on estimating the value that parents place on school quality by calculating the premium paid for houses located in areas with better schools (e.g., [Black \(1999\)](#) and [Kane et al. \(2006\)](#)).

Several authors study the determinants of school choice in the Chilean context.⁶ For instance, [Gallego and Hernando \(2008\)](#), using a semi-structural approach, find results that suggest that the school choice implemented in Chile increased overall student welfare, but they also find that there is a lot of heterogeneity in the size and even the sign of the welfare change. Along the same lines, [Chumacero et al. \(2011\)](#), using a database that accurately estimates the distance between the household and school, find that both quality and distance are highly valued by households. In a novel paper, [Neilson \(2013\)](#) studies the effects of targeted school vouchers on the outcomes of poor children in Chile;⁷ his findings suggest that this program effectively raised competition in poor neighborhoods, pushing schools to improve their academic quality. Finally, [Mark et al. \(2006\)](#) study how parents construct their school choice sets and comparing this to what they say they are seeking in choosing schools. Their results indicate that parental decisions are influenced by demographics.

Furthermore, this paper is related to the literature that models individuals' economic deci-

⁵See for example [Alderman et al. \(2001\)](#), and [Bast and Walberg \(2004\)](#). In a very interesting paper, [Epple et al. \(2018\)](#) estimate a model of parents school choice, providing evidence that students sort into schools with a higher proportion of other students similar to them. Using the estimated model, they discuss what is the optimal policy for closing schools in the context of declining enrollments.

⁶We describe Chile's school choice policies in the next section.

⁷This paper studies the impact of the SEP policy (the exogenous variation used in this paper). Other papers that do so are [Aguirre \(2018\)](#) and [Sánchez \(2018\)](#).

sions incorporating bounded rationality. In general, this literature follows the idea that, as [Simon \(1986\)](#) points out, cognitive effort is a scarce resource, and the knowledge and computational power of the decision-maker are always limited. In an interesting paper, which is one of the few papers that perform a structural estimation with bounded rationality, [Houser et al. \(2004\)](#) develop a Bayesian procedure for classification of subjects into decision rule types in choice experiments, finding that, in a very difficult dynamic problem, more than a third of the experimental subjects followed a rule very close to the optimal (expected wealth maximizing) rule.⁸

Because school choice is a complex task, which involves gathering and processing information, different authors have studied the presence of bounded rationality in that context. For instance, [Schneider et al. \(1998\)](#) find that, on average, low-income parents have very little accurate information about objective conditions in the schools.⁹ However, even though levels of objective information held by parents are low, their actual choice of schools reflects their preferences in education. Along the same lines, [Azmat and Garcia-Montalvo \(2012\)](#) conclude that, as well as parents' education, information gathering and information processing are important determinants for the quality of school choice.

Finally, this paper is also related to the literature that combines sources of exogenous variation and the empirical methods that use this variation to obtain causal estimates, as experimental data, regression discontinuity, and so on; with the structural estimation approach. In the labor economic literature, the seminal papers are [Wolpin and Todd \(2006\)](#) and [Attanasio et al. \(2012\)](#), where only the latter uses the exogenous variation (RCT) to estimate the model, while the former uses that variation to validate the model.¹⁰ More recent and interesting papers along this line, are [Allende et al. \(2019\)](#) and [Fu and Gregory \(2018\)](#).

3. Chile's School System

In 1981, the Chilean military government created a voucher market in the educational system, which was part of a broader reform that also included the decentralization of public

⁸This paper is also related to the literature that attempts to infer agents' information using observed choices, such as: [Carneiro et al. \(2003\)](#); [Cunha et al. \(2005\)](#); and [Navarro \(2011\)](#).

⁹The same is found by [Henig \(1996\)](#).

¹⁰In the IO literature this combination has been exploited much before than in the labor case (with semi structural approaches), see for example the influential work of [Berry et al. \(1995\)](#).

schools (which were transferred to municipalities) and the introduction of flexibility in teachers' contracts.¹¹ This reform transformed the way schools were funded by the government, establishing a system where private and public schools were paid per student, with a flat voucher, on the basis of attendance.

Since then, the allocation of public resources has been mainly determined by parents' decisions. However, in practice, this decision has had several restrictions: schools can select students based on their previous performance, tests, and the characteristics of their parents (e.g., marital status and religion). On top of that, since 1994, when a co-payment law was passed, schools that are eligible for public funding can also charge a tuition fee; in that case, depending on the amount charged, there is a discount to the school's subsidy.¹²

These reforms consolidated a system of mixed provision of education, with three types of schools; municipal (public), private subsidized (voucher-private), and entirely private (non voucher-private). The first two receive most of their funds from state vouchers, and, since 1994, privately subsidized schools may additionally charge a tuition fee. In 2016, over 90% of the Chilean students received funding via vouchers.¹³

As stated above, in 2008 the implementation the Ley de Subvención Escolar Preferencial (SEP) increased the voucher for the socioeconomically vulnerable students by 60%, on the top of the regular voucher. The explicit aim of this additional funding was to increase quality in schools attended by the most vulnerable students. There were two components of the SEP subsidy: i) each school receives a fixed amount for each vulnerable student and ii) each school receives an additional subsidy for each vulnerable student, if the fraction of vulnerable students in the school is "high" (the law established brackets according to these shares, with higher per-student subsidies for higher shares). Furthermore, this new law not only transferred significantly more resources to schools for each socioeconomically vulnerable matriculated, it also eliminated the out-of-pocket tuition fees for those families. In practice,

¹¹For a summary of these reforms, see [Gauri \(1999\)](#) and [Mizala and Romaguera \(2000\)](#).

¹²[Epple and Romano \(2008\)](#) emphasize the consequences of this selection mechanism for the outcomes of an educational voucher system. Structural reforms have been approved in January 2015 by the Chilean parliament. After a transition period, these reforms will imply that private-voucher schools will not be allowed to charge tuition, to be for profit, and to select students; the entry of new private-voucher schools has been restricted. Because this paper considers the period between 2008 and 2015, our results are not affected by these reforms.

¹³Source: Ministry of Education, Chile.

this extra voucher applied to almost half of students at most schools. As discussed, the exogenous variation generated by this policy is the key element in the identification strategy of this paper.

In order to guide parents' decisions and to measure the student learning process, a new testing system, SIMCE, came into existence in 1988. The SIMCE is an annual nationwide standardized test. Its results have been public information at school level since 1995, publicized in part by listings in major newspapers of individual schools' performance. The government also uses SIMCE scores to allocate resources.¹⁴

More than 30 years after the reform, there are several clear stylized facts. First, Chile is one of the most unequal OCDE countries in terms of students results in standardized tests (although the situation is even worst in other latin american countries). Second, the magnitude of socioeconomic school segregation is very high (higher than the geographical segregation), and has increased slightly over the last decade (Valenzuela et al. (2014)). Third, despite important increases in the public budget allocated to education, Chile's performance is relatively poor when compared with similar countries (Chumacero et al. (2011)). Finally, there has been a massive migration from public to private schools, which was accelerated after passage of the co-payment law (See Larrañaga (2004)). Indeed, the student fraction in the public system went from 78%, in 1981, to 38% in 2012.

4. Data and Basic Statistics

4.1. Sources of data

The main source of information in this paper is a new survey of the Ministry of Education of Chile, the *Encuesta de Eleccion Escolar* (EEE), which was designed to study parents' school choice in the context of the implementation of a new admission system to primary and secondary schools.¹⁵ This survey was carried out in 2015, one year before the implementation

¹⁴Meckes and Carrasco (2010) describe SIMCE's main features, purposes, institutional framework, and strategies for communicating results.

¹⁵This new admission system started in 2016 in the Region of Magallanes (the extreme south of Chile). It has two main differences with respect to the previous system. First, schools cannot longer select students based on their characteristics or the characteristics of their families. Second, students are allocated in a centralized fashion, using the deferred-acceptance algorithm (see Abdulkadirolu et al. (2005)). This allocation process uses the parents' preferences as reported by them (*i.e.*, a ranking of the schools they would like to send their child to) and seeks to maximize parents' utilities, without creating incentives for their strategic behavior.

of the new system, and it has information about the school attended by the students, their parents’ education, and the reasons why parents chose that school. Moreover, the key aspect of the data for this paper is that it includes measures of parents ability to process information about schools’ test scores. In particular, parents in the survey are asked to rank the test scores of three schools they know and like, in most of the cases including the school that is actually attended by their child, relative to other schools in their neighbourhood. This information is critical for us to determine whether parents are able to rank schools properly in terms of their average test scores in the Chilean standardized test (the SIMCE).

To create the sample for the EEE, the Chilean territory was divided into educational markets. Among 104 markets that were identified, 30 were selected based on the following pre-specified requirements: to have at least 20 schools, to have at least 500 vulnerable students, and to have at least one private-voucher school.¹⁶ The full sample considers households with children who are in school, as well as those who are about to enter. However, because ours is a model of the first school choice, we need data on the actual choice for our estimation. Thus, we restrict the sample to only those households whose child is already attending school.

We complement the information from the EEE with administrative information from the Ministry of Education of Chile. In order to characterize schools, we use SIMCE test scores and the information collected from SIMCE parents’ surveys.¹⁷ Test scores are used to estimate schools’ quality every year. The surveys also include questions about school tuition fees and whether the school considered some of the following elements in the admission process: a student test, parents’ marital status (and whether they had a religious wedding), and a general “other” category to account for any other information considered in the admission. We also like to categorize schools in terms of socioeconomic status (SES). To achieve this, we aggregate information on the schooling of parents with children attending a given school. Furthermore, from administrative data of the Ministry of Education, we collect information about each school’s religious affiliation, their type of administration (public, voucher-private, and non voucher-private), and whether it is located in an urban or rural area.

¹⁶To select those 30, there was a stratification by size of the market, percentage of vulnerable students, fraction of total enrollment attending private-voucher schools, and the timing of the geographical implementation of the new admission system (the different regions enter the new system sequentially over a period of three years). Moreover, there was an over-representation of households living in markets where the new system is implemented earlier.

¹⁷We use the 4th grade SIMCE from 2006 to 2015.

As we have mentioned, we use an instrument for school quality in our estimation. To build this instrument, we use administrative data on the monetary supplement to the regular value of the voucher for each vulnerable attending each school (which varies over time), given by the 2008 law we described in section 3 and have labeled SEP. We describe this instrument in detail below.

4.2. Definition of variables

We now list the variables we use in our estimation. We group these variables according to the different equations that enter in the model we present in section 6. For each of these groups, we include a description (name) and the notation we use for each of them in the model.

School level variables (Y) This vector includes dummy variables for the type of school (public, private with voucher, private); dummy variables for the socioeconomic status of the school; a dummy variable indicating whether it is a rural or urban school; and the schools' tuition fee in 2007. This is the average fee reported by parents when they answer the SIMCE survey. This vector also includes variables that characterize the school before 2007, the year when the SEP was introduced. In particular, it has a polynomial on the share of students that attended the school on 2006 and a polynomial on the fraction of voucher students that attended the school on 2006.

Variable relating school and household characteristics (X) We include a dummy variable that takes the value of one if it is a religious school and the family reports to care about the religious affiliation of schools when making choices.

Household characteristics for the information type probabilities (W) We aggregate the information from the EEE in one single measure that aims to capture the degree to which households are able to rank schools they know. The EEE first asks households about a group of 10 schools in their neighbourhood. Then, households are asked to name other schools they know. Among this expanded set of schools, they are then asked to name the three schools they would prefer for their child, if they had to choose one to enroll him/her.

¹⁸ For each of these three schools, they are asked to say in which quintile of the distribution of schools of their neighborhood they think this school is. We then calculate the fraction of times (out of three) they are correct. Throughout this paper, we refer to this measure as the *success index*.

Household characteristics shifting preferences for quality (H) As mentioned before, following the literature, we allow preferences for quality to vary according to the mother’s education. We categorize mother’s education in three groups: elementary, secondary, and any tertiary education (technical, college, post graduate).

Variables related to schools’ selection processes (M) The survey that is collected together with the SIMCE test asks parents whether their child’s school selects students based on religion, whether they are married, a test, or any other criteria. We aggregate these responses at the school level to build a measure of each school’s selectivity based on each of these criteria. We then interact these measures with relevant household characteristics: the religious selectivity with a dummy indicating if the parents say they are religious; the marital selectivity with a dummy indicating if the parents are married; and each of the other two criteria interacted with the mothers education level, which we use as a proxy for the household’s SES.

Quality (Q) We use SIMCE math scores as a measure of “true” quality. Although a school’s SIMCE score is really a function of its quality as well as the non-random pool of abilities of students enrolled, we choose to use scores as measures of quality because these are precisely the information available to parents to help them choose schools (see section 3) and the measure used to ask parents to rank schools in the data we use to build the measures in W (the success index). In the model and estimation, for reasons that will become evident in section 6, these school qualities are demeaned with respect to the average quality in each household’s choice set.

¹⁸The specific question is: “If we ordered all of the schools in your neighbourhood according to their SIMCE score, in what group or quintile do you think [School i] would be in comparison to other elementary schools in your neighborhood?” (Original in Spanish: Si ordenamos todos los establecimientos de su barrio por puntaje SIMCE ¿en qué grupo o quintil cree que se encuentra el [Establecimiento i] al compararlo con los establecimientos de básica de su barrio?)

Instrumental variable for quality (Z) As we have described, our instrumental variable exploits variation in schools' funding due to the implementation of the SEP, which introduced additional funding for schools based on the enrollment of socioeconomically vulnerable students (vulnerable students, henceforth). There are two components of the SEP subsidy: i) each school receives a fixed amount for each vulnerable student and ii) each school might receive an additional subsidy for each vulnerable student, depending on where the fraction of vulnerable students in the school lies within brackets established by the government. For example, a school with few vulnerable students may receive $\$s_1$ dollars for each of these students, while another one with a share of vulnerable students higher than a specified threshold would receive $\$s_1 + \s_2 dollars for each of its vulnerable students. Because the additional funds received by a school with an additional vulnerable student are not necessarily equal to the marginal cost of providing the education services to that student, we argue that this additional funding may affect quality.

However, the funding from the SEP is endogenous for the same reasons that quality is endogenous: it is, by definition, related to the school choices of parents. Nevertheless, using the rule that assigns funds based on socioeconomic composition of schools, we can build the funding that each school would have received had the socioeconomic composition remained as it was before the SEP was implemented. Particularly, we use the socioeconomic composition of schools in the year immediately before our estimation period (also before the SEP was implemented), which we will refer to as the baseline year.

Formally, our instrument z_{jt} (for school j in year t) is:

$$z_{jt} = (s_{1,t} + s_{2,t}(N_{j,0})) * N_{j,0} \quad (1)$$

Where, $N_{j,0}$ is the share of vulnerable students in school j in the baseline year; $s_{1,t}$ is the subsidy allocated for each vulnerable student in year t ; and $s_{2,t}$ is the additional per-student subsidy assigned in year t , which is a function of the same share of vulnerable students, $N_{j,0}$.

4.3. Definition of choice sets

For the estimation of the structural model, we need to define the set of potential schools a particular student might attend (which we do not observe). Using the entire pool of schools

as the choice set for any student has various problems. On the one hand, this would make the size of the choice set infeasibly large for estimation. On the other hand, there are schools that are not really considered as possible options by parents. Some of these are for obvious reasons (e.g. the school is in a different city) but other are more subjective. To address both of these issues, we define the choice set for each student (Λ_i) in the following way:

1. We force choice sets to include between 10 and 50 schools.
2. The school actually attended by student i and the other two schools ranked in the survey by his/her parents, are always in his/her choice set.
3. For student i , we begin by including all of the schools that are chosen by any other student in the same municipality and same socio-economic background (defined using the categories of the student's mother's education).
4. If the set of schools defined using the previous criteria includes less than 10 schools, we do not restrict by socio economic background in step 3.
5. If the set of schools includes more than 50 schools, we eliminate schools at random (always keeping the true school and the other two schools ranked in the survey in the choice set).

Defining the choice sets in this way allows us to use the available information (observed parents choices) to define a crucial component of our model that is, by definition, unobservable. The choice sets we define in this way will reflect all binding restrictions faced by parents, which may result from their location or even deeper causes, such as socioeconomic segregation. Therefore, having a large number of families is crucial. Under these circumstances, if we do not observe any family belonging to a specific group (as defined above) choosing a particular school, we interpret this observation as meaning that it is not feasible for that group of families to attend that school. Intuitively, this is an identification at infinity type of argument.

4.4. Sample

Our final sample has 1894 families. These are families with a child who started elementary school between 2008 and 2015 and were interviewed in the EEE. Moreover, the EEE asks

about only one child per household. Therefore, by design of the EEE, the children in our sample tend to be the youngest in their households (EEE selects the youngest child when there is more than one, to focus on children with more recent school choices).

Table 1 shows descriptive statistics for our final sample. On average, households categorize one out of three schools correctly when asked to compare them to the relevant distribution. This, on itself, suggests relevant limitations in terms of how well parents can compare schools based on simple measures of academic quality. These statistics also show that there is a non/trivial amount of selection carried on by schools, both based on knowledge or abilities and also on family characteristics. The majority of students attend private voucher schools, although a sizable fraction still chooses public schools. Finally, the positive difference between the demeaned quality and the standardized score (Math Score Std.) suggests that, by using the demeaned quality, we are adjusting to different levels of quality available to different households, but that this adjustment is not extremely restrictive.

As mentioned, the sample of students in the EEE is a non/random sub-sample of students entering elementary school between 2008 and 2015. We compare our sample to the whole population of students in Table 2. As it can be seen, children in our sample attend schools that have a slightly higher academic quality, are more likely to be public, and show a higher degree of selectivity. Families attending these schools tend to be of a better socio-economic background. Despite most of these differences being statistically significant, the magnitudes are not very large. Thus, even though we work with a sample that is clearly selected, it does not display substantial differences with respect to the entire population.

Table 1: ESTIMATION SAMPLE' SUMMARY STATISTICS

	Mean	Std. Dev.	Min	Max
Family characteristics:				
Success Index	.305	.319	0	1
Family Religion (*)	.126	.332	0	1
Characteristics of the schools attended by families:				
School Quality	.113	.447	-1.366	1.596
Math Score Std.	.01	.483	-1.464	1.509
Sel. by civil marriage (*)	.045	.068	0	.559
Sel. by Knowledge (*)	.288	.311	0	1
Sel. by Others aspects (*)	.189	.111	0	.664
Sel. by Religion (*)	.106	.244	0	.985
Rural (*)	.04	.196	0	1
Public (*)	.431	.495	0	1
Private with voucher (*)	.563	.496	0	1
Private without voucher (*)	.005	.072	0	1
SES of School 1 (*)	.07	.256	0	1
SES of School 2 (*)	.302	.459	0	1
SES of School 3 (*)	.418	.493	0	1
SES of School 4 (*)	.2	.4	0	1
SES of School 5 (*)	.01	.097	0	1
School Religion (*)	.511	.5	0	1
Tuition Fee 2007	8862	13810	0	108,660
SEP Instrument	56,228	61,511	0	332,122
Sample size	1,894			

Note: This table presents the characteristics of the families included in the estimation sample and the characteristics of the schools attended by the. (*) is for fractions.

Table 2: SCHOOLS' CHARACTERISTICS: ESTIMATION SAMPLE VERSUS POPULATION

	Population	Estimation Sample	Difference	S.E.	P-Value
Math Score Std.	-0.059	0.010	-0.069	0.011	0.000
Sel. by civil marriage (*)	0.036	0.045	-0.009	0.002	0.000
Sel. by Knowledge (*)	0.267	0.288	-0.021	0.007	0.004
Sel. by Others aspects (*)	0.183	0.189	-0.005	0.003	0.044
Sel. by Religion (*)	0.081	0.106	-0.026	0.006	0.000
Rural (*)	0.117	0.040	0.077	0.005	0.000
Public (*)	0.450	0.431	0.018	0.011	0.110
Private with voucher (*)	0.538	0.563	-0.025	0.011	0.028
Private without voucher (*)	0.012	0.005	0.007	0.002	0.000
SES of School 1 (*)	0.106	0.070	0.036	0.006	0.000
SES of School 2 (*)	0.333	0.302	0.031	0.011	0.003
SES of School 3 (*)	0.373	0.418	-0.045	0.011	0.000
SES of School 4 (*)	0.168	0.200	-0.032	0.009	0.001
SES of School 5 (*)	0.020	0.010	0.010	0.002	0.000
School Religion (*)	0.469	0.511	-0.042	0.012	0.000
Fee 2007	10,469	8,862	1,607	320	0.000
SEP Instrument	73,342	56,228	17,114	1,428	0.000
Number of obs.	136,424	1,894			

Note: This table presents the characteristics of the schools attended by parents, both for the population and for our estimation sample, where *population* uses the information of the 2015 SIMCE (4th grade). (*) is for fractions.

5. Success Index and Reduced Form Estimates

In this section, we present a set of descriptive and non-structural estimates that summarize key aspects of the data. First, we study the characteristics of the families that are more able to rank schools properly according to our measure, the success index. Second, we present evidence that this measure is not biased in ways that would challenge the identification. Third, we show, in a reduced form way, estimates that suggest that school quality is more relevant in the choices made by wealthier families and by parents who are able to rank school performance properly. The structural model we present in section 6 will specify the assumptions under which we draw inference from these features of the data.

5.1. Success Index

To study what are the characteristics of the families that are more able to rank schools properly, we simply regress the success index on a series of family characteristics that play a role in our model. As Table 3 shows, more educated families are monotonically more able to rank schools properly. Married families and families that profess a religion also have a higher probability to rank schools properly. However, in these last two cases, the estimated parameters are not statistically significant. All of these results show how our measure of the ability to compare schools is correlated to other observable characteristics (as expected). In particular, the correlation with education will play an important role in the results and interpretation of the counterfactual exercises we conduct in section 9.

We now study two possible sources of bias in our measure of parents ability to rank schools in terms of their standardized test performance. In the first place, it is possible that parents whose children have been attending primary schools for longer (*i.e.*, they are in higher grades), are more able to rank schools. Since parents are asked to rank schools when their children are attending different grades, the existence of this relationship could bias our measure. In the second place, it is possible that parents are more able to properly rank the school that is being attended by their child. To the extent that not all parents include this school in the set of three schools they rank, this could also bias our measure of parents ability to rank schools.

In Table 4 we present the estimated models to address these concerns. These are linear

Table 3: SUCCESS INDEX AND DEMOGRAPHIC CHARACTERISTICS

Dependent Variable:	Success Index
	(1)
Parent with completed secondary education	0.040 (0.018)
Parent with completed tertiary education	0.097 (0.018)
Dummy for Religion	0.020 (0.021)
Dummy for Married	0.013 (0.015)
Observations	1,894
R-squared	0.016

Note: In this table an OLS estimation is reported, where there is one data point for each individual considered in our structural estimation. The information about family education is built given the answers given by the surveyed person, in this case the omitted category is *primary education completed*, which includes those who attended secondary school without completing that level.

probability models where, for each household, we have as many observations as schools are ranked by parents. The dependent variable takes the value of one if parents are able to rank the school properly and zero otherwise. In order to test the existence of these sources of bias in the success index, we consider two independent variables of interest. First, we look at a dummy variable that takes the value of one if the school is the one attended by the student and zero otherwise. Second, we include dummies that capture how long the students have been in the school system (essentially, their grade). We include three different specifications. In column (1), we run an OLS model which only considers these two independent variables of interest. We then run the OLS model including family characteristics as control variables and show these results in column (2). Finally, we run an OLS model including individual fixed effects. Because the only variable that is different for different observations of the same individual is the dummy for the elected school, all the other variables drop out in this specification.

The results of these estimations support the index that we use to measure parents ability to rank schools. None of the specifications show evidence that parents are more likely to rank correctly the school that is attended, relative to other schools they include in their ranking. With respect to the number of years in the school system, although the dummy

variable indicating 2, 3, or 4 years is statistically significant at the 10% level, we are not concerned about its estimated magnitude (three percentage point difference). Furthermore, there is no evidence that having been in the school system for 5, 6, or 7 years (relative to only one) makes parents more likely to rank schools appropriately.

Table 4: PROBABILITY OF RAKING EACH SCHOOL PROPERLY

	OLS		FE ID
	(1)	(2)	(3)
Dummy for School attended	-0.009 (0.018)	-0.011 (0.017)	-0.003 (0.021)
Dummy for 2, 3, and 4 years attending primary school	-0.036 (0.019)	-0.031 (0.019)	
Dummy for 5, 6, and 7 years attending primary school	-0.006 (0.021)	0.003 (0.021)	
Parent with completed secondary education		0.050 (0.020)	
Parent with completed tertiary education		0.097 (0.020)	
Dummy for Religion		0.013 (0.022)	
Dummy for Married		0.006 (0.016)	
Observations	3629	3629	3399
Adjusted R-squared	0.001	0.008	0.432

Note: In this table an OLS estimation is reported, where there is one data point for each school ranked by parents in the survey, in some cases there are 3 schools ranked, in others two or one. The omitted category is *primary education completed*, which includes those who attended secondary school without completing that level. In the last specification (column 3) fixed effect at individual level is considered. Robust standard errors are in parentheses, and they are clustered at individual level. ***, ** and * indicate statistical significance at the 1%, 5% and 10%, respectively.

5.2. Reduced form estimates

To motivate the estimation of the structural model we describe in section 6, we estimate the reduced form regressions shown in Table 5. As we described in section 4, we have defined a choice set for each household. These choice sets, however, may have different sizes for different households. Thus, we first define a “small choice set” of 10 schools for each of the students in our sample ($10 * 1,894$ observations). This “small choice set” is composed by the school attended by the student, the other two schools ranked by parents and seven other ones chosen at random from the actual choice set as defined in section 4. We choose 10 schools for these “small choice sets” as this is the size of the smallest choice set in the sample, based

on how we define them. This means that the “small choice set” used to estimate the reduced form specification is the same as the actual choice set for some households and different for other households.

Our dependent variable is a dummy variable that takes the value of one if the school was chosen by the student, and zero otherwise. We regress this variable against the school’s academic quality and its interaction with the success index and the education level of the person who answered the survey (a parent in most of the cases). In the structural model, these variables are the variables related to heterogeneity in the ability to tell schools apart and preferences, respectively. We also control for all the other variables that appear in the structural model.

In the first two columns, we show estimates from OLS and IV models that include the demeaned quality as the main explanatory variable. The results show that academic quality has a small but positive effect on school choice (the IV estimate is not statistically significant, but this appears to be a problem of precision, not bias). However, when we include the interaction of quality with our success index, which measures how well parents can distinguish between schools with different academic qualities, the relevance of quality becomes much more predominant. Moreover, quality becomes irrelevant for those households who cannot distinguish at all between different quality schools. We obtain similar results when we include the interactions with family education in columns 5 and 6. Quality appears to have no effect for low educated households (no more than completed primary education) and the effect increases with education. These results remain when we include both sets of interactions in columns 7 and 8.

The Kleibergen-Paap Wald F (KP) statistic of the first stage for the IV estimate can be found in the last row of Table 5. As this table shows, our instrument has enough statistical power, namely its F-statistic is 31.36, far above the value of ten or any critical value for test of weak instruments.¹⁹ Moreover, as Table 6 shows, the sign in the first stage regression of the impact of the instrument on school quality is as expected, in the sense that those schools

¹⁹The KP statistic decreases for the estimations in columns 4,6, and 8 of Table 5. However, this is a consequence of the fact that, for these reduced form estimations, we need to instrument for all the interactions as well (i.e. we have a larger number of first stage equations). In the structural estimation, as it will become evident in section 7, there is only one first stage equation. Thus, the relevant statistic to assess relevance of our instrumental variable is the KP presented in Table 6.

with low socioeconomic composition before the law was passed, were the ones whose quality increased the most after the implementation of the SEP.

Table 5: REDUCED FORM ESTIMATION FOR SCHOOL CHOICE

	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Quality	0.021 (0.006)	0.050 (0.108)	0.001 (0.007)	-0.014 (0.118)	0.008 (0.009)	-0.123 (0.151)	-0.011 (0.009)	-0.171 (0.158)
Quality $\times$ Success Index			0.067 (0.016)	0.216 (0.083)			0.066 (0.016)	0.200 (0.083)
Quality $\times$ SES 2 of Family					0.013 (0.010)	0.258 (0.083)	0.013 (0.010)	0.247 (0.082)
Quality $\times$ SES 3 of Family					0.025 (0.011)	0.218 (0.083)	0.021 (0.012)	0.199 (0.083)
Sel. in Math ability	-0.021 (0.070)	-0.021 (0.070)	-0.021 (0.070)	-0.020 (0.070)	-0.019 (0.069)	0.003 (0.071)	-0.019 (0.069)	0.002 (0.072)
Sel. in Knowledge	-0.037 (0.027)	-0.048 (0.048)	-0.035 (0.027)	-0.045 (0.049)	-0.025 (0.027)	0.094 (0.085)	-0.025 (0.027)	0.089 (0.085)
Sel. in Others aspects	0.055 (0.059)	0.037 (0.085)	0.052 (0.059)	0.026 (0.085)	0.047 (0.059)	-0.050 (0.077)	0.045 (0.059)	-0.054 (0.076)
Sel. MATH $\times$ SES 2 of Family	-0.103 (0.094)	-0.100 (0.093)	-0.105 (0.094)	-0.105 (0.095)	-0.104 (0.094)	-0.140 (0.095)	-0.106 (0.094)	-0.144 (0.097)
Sel. MATH $\times$ SES 3 of Family	0.124 (0.096)	0.124 (0.095)	0.119 (0.096)	0.107 (0.097)	0.119 (0.095)	0.090 (0.095)	0.115 (0.095)	0.078 (0.097)
Sel. Knowledge $\times$ SES 2 of Family	0.033 (0.032)	0.033 (0.032)	0.030 (0.032)	0.025 (0.033)	0.022 (0.033)	-0.171 (0.073)	0.020 (0.033)	-0.169 (0.074)
Sel. Knowledge $\times$ SES 3 of Family	0.096 (0.032)	0.098 (0.031)	0.091 (0.031)	0.081 (0.032)	0.078 (0.032)	-0.070 (0.072)	0.076 (0.032)	-0.072 (0.072)
Sel. Others $\times$ SES 5 of Family	-0.007 (0.058)	0.000 (0.068)	-0.007 (0.058)	0.001 (0.068)	-0.001 (0.058)	0.133 (0.074)	-0.001 (0.058)	0.128 (0.075)
Sel. Others $\times$ SES 6 of Family	-0.200 (0.057)	-0.178 (0.104)	-0.194 (0.057)	-0.155 (0.102)	-0.186 (0.057)	-0.060 (0.089)	-0.182 (0.057)	-0.050 (0.088)
SES of School 2	0.006 (0.007)	0.004 (0.009)	0.006 (0.007)	0.006 (0.010)	0.006 (0.007)	0.011 (0.011)	0.007 (0.007)	0.012 (0.011)
SES of School 3	0.036 (0.013)	0.030 (0.030)	0.037 (0.013)	0.033 (0.030)	0.037 (0.013)	0.038 (0.032)	0.038 (0.013)	0.040 (0.032)
SES of School 4	0.029 (0.021)	0.016 (0.055)	0.029 (0.021)	0.018 (0.056)	0.029 (0.021)	0.019 (0.056)	0.030 (0.021)	0.020 (0.056)
SES of School 5	0.008 (0.037)	-0.010 (0.080)	0.011 (0.037)	-0.002 (0.082)	0.005 (0.037)	-0.024 (0.077)	0.009 (0.037)	-0.015 (0.079)
Dummy for Rural School	-0.070 (0.007)	-0.072 (0.012)	-0.070 (0.007)	-0.073 (0.012)	-0.070 (0.007)	-0.073 (0.012)	-0.070 (0.007)	-0.073 (0.012)
School Dependence: Subsidized School	0.018 (0.021)	0.019 (0.021)	0.018 (0.021)	0.019 (0.021)	0.018 (0.021)	0.020 (0.021)	0.018 (0.021)	0.021 (0.021)
School Dependence: Privates School	0.012 (0.039)	0.016 (0.041)	0.012 (0.039)	0.014 (0.043)	0.012 (0.039)	0.017 (0.043)	0.012 (0.039)	0.016 (0.044)
School Religion $\times$ Family Religion	0.026 (0.011)	0.027 (0.012)	0.026 (0.011)	0.026 (0.012)	0.027 (0.011)	0.030 (0.012)	0.026 (0.011)	0.028 (0.012)
Fee 2007	-0.001 (0.000)	-0.001 (0.001)	-0.001 (0.000)	-0.001 (0.001)	-0.001 (0.000)	-0.001 (0.001)	-0.001 (0.000)	-0.001 (0.001)
SEP Fraction 2008	0.017 (0.089)	0.013 (0.092)	0.018 (0.089)	0.016 (0.092)	0.018 (0.090)	0.036 (0.093)	0.019 (0.090)	0.037 (0.094)
Enrollment Fraction 2006	0.935 (0.235)	0.891 (0.323)	0.935 (0.235)	0.887 (0.322)	0.936 (0.235)	0.916 (0.324)	0.936 (0.235)	0.912 (0.323)
Enrollment Fraction 2006 Square	-1.262 (0.279)	-1.210 (0.378)	-1.264 (0.279)	-1.211 (0.378)	-1.264 (0.279)	-1.237 (0.381)	-1.265 (0.279)	-1.238 (0.381)
SEP Fraction 2008 $\times$ Private School Dependence	-0.035 (0.044)	-0.039 (0.046)	-0.036 (0.044)	-0.040 (0.046)	-0.036 (0.044)	-0.046 (0.046)	-0.036 (0.044)	-0.047 (0.046)
Number of Obs.	18940	18940	18940	18940	18940	18940	18940	18940
R-squared	0.041	0.039	0.042	0.034	0.041	0.016	0.043	0.013
Kleibergen-Paap Wald F statistic		26.818		14.355		8.960		6.913
Cragg-Donald Wald F		159.911		76.388		50.507		36.911

Note: This table reports the second stage estimation (it is a linear probability model). The instrumental variable is the funding that each school would have received had the socioeconomic composition remained as it was before the SEP was implemented, described in equation 1. There are ten schools for each individual, where only one has the value of one, and the other nine, the value of zero. Robust standard errors are in parentheses, and they are clustered at individual level.

Table 6: FIRST STAGE REGRESSION

Z : SEP Instrument	0.002 (0.000)
Sel. in Math ability	0.003 (0.173)
Sel. in Knowledge	0.399 (0.057)
Sel. in Others aspects	0.625 (0.135)
Sel. MATH $\times$ SES 2 of Family	-0.097 (0.162)
Sel. MATH $\times$ SES 3 of Family	-0.006 (0.169)
Sel. Knowledge $\times$ SES 2 of Family	-0.003 (0.046)
Sel. Knowledge $\times$ SES 3 of Family	-0.035 (0.048)
Sel. Others $\times$ SES 5 of Family	-0.291 (0.102)
Sel. Others $\times$ SES 6 of Family	-0.785 (0.103)
SES of School 2	0.052 (0.027)
SES of School 3	0.217 (0.033)
SES of School 4	0.440 (0.043)
SES of School 5	0.631 (0.076)
Dummy for Rural School	0.096 (0.025)
School Dependence: Subsidized School	-0.013 (0.042)
School Dependence: Privates School	-0.112 (0.099)
School Religion $\times$ Family Religion	-0.037 (0.014)
Fee 2007	0.003 (0.001)
SEP Fraction 2008	-0.310 (0.192)
Enrollment Fraction 2006	1.442 (0.233)
Enrollment Fraction 2006 Square	-1.686 (0.297)
SEP Fraction 2008 $\times$ Private School Dependence	0.117 (0.134)
Number of Obs.	18,940
Kleibergen-Paap Wald F statistic	26.818

Note: This table reports the first stage estimation (that corresponds to column (2) of Table 5). The instrumental variable (Z: SEP instrument) is the funding that each school would have received had the socioeconomic composition remained as it was before the SEP was implemented, described in equation 1. There are ten data point for each individual for each individual. Robust standard errors are in parentheses, and they are clustered at individual level.

6. Model

We consider a simple static model in which each household $i \in \{1, 2, \dots, I\}$ with a child that is about to start elementary school decides among its possible elementary school alternatives.²⁰ The educational market is composed of J schools. However, each household's choice is restricted in two ways. First, each household i has a specific choice set $\Lambda_i \subseteq \{1, 2, \dots, J\}$, with cardinality denoted by $S(\Lambda_i)$. Second, each school $j \in \Lambda_i$ may or may not admit the student i based on a series of rules that we describe below.

6.1. Households' Utility

The flow utility of household i when their child attends school j is given by:

$$u_{ij} = \theta_j + X_{ij}\beta^X + \beta_i^q \tilde{q}_{ij} + \varepsilon_{ij} \quad (2)$$

There are three key elements in this flow utility equation, besides the error term ε_{ij} . First, θ_j captures how much school j attracts households in general, for reasons other than academic quality. Ideally, we would estimate all the θ_j 's non-parametrically. However, this is unfeasible, as it would imply estimating thousands of parameters, and we would need a large enough number of households with each school in their choice set to do so. Unfortunately, the sample we use does not allow us to follow this route. As an alternative, we specify θ_j to be a parametric function of observed characteristics of the school (Y):

$$\theta_j = Y_j \beta^Y \quad (3)$$

where the vector Y contains information aimed at capturing the propensity of households to pick school j in general. Specifically, it includes a polynomial on the share of students that attend school j , a polynomial on the fraction of voucher students that attend school j , whether the school is located in a rural or urban area, the school's tuition fee, dummy variables indicating whether the school is public, private with voucher or private without voucher, and interactions between these variables. Naturally, many of these measures are equilibrium outcomes. To address this issue, we include these measures (market shares and

²⁰We treat the household and student in that household as a single unit.

prices) as measured before our sample period.

The second term, $X_{ij}\beta^X$, represents characteristics that are specific to the match between household and school (e.g. if the school is religious and the household cares about religion), thus capturing how household i might be especially attracted to school j . Finally, the term we are most interested in, $\beta_i^q \tilde{q}_{ij}$, represents how household i cares about school j 's academic quality.

We would like to highlight two features of this last term. First, note that the parameter multiplying the quality, β_i^q , may be heterogeneous across households. Specifically, we define $\beta_i^q = \exp(H_i\alpha)$. In the empirical implementation, H is a vector of indicators of household's education level, which serves as a proxy for socio-economic background.²¹ ²² We choose to allow this heterogeneity according to parental education based on differences that have been previously found in the literature (see, for example, [Hastings and Weinstein \(2008\)](#)). Second, quality is indexed both by i and j . This represents the fact that this is the quality of school j as perceived by household i . We formalize this notion below.

6.2. Perceived quality

We are interested in potential limitations to the way households (parents) perceive quality. Thus, we allow households to be of different *types*, depending on how well they can distinguish among the quality of schools available to them. Let $k \in \{1, 2, \dots, K\}$ denote each of these types. We define the perceived quality for household i if it is of type k as a convex combination of the true quality q_j and the mean quality of schools in household i 's choice set $\bar{q}_i$:

$$\tilde{q}_{ijk} = \tau_k q_j + (1 - \tau_k) \bar{q}_i \quad (4)$$

with $\tau_k \in (0, 1) \forall k$.

If τ were equal to one, the perceived quality would be the true quality. On the other hand, a household with $\tau = 0$ would perceive all of the schools in its choice set as having the same quality (equal to the mean true quality).

Replacing the perceived quality in the utility function and re-arranging some terms we

²¹Household's education is based on the education of the person who answers the EEE survey, usually one of the parents

²²We will need to normalize $\beta_q = 1$ for one of these education groups. We will come back to this point later.

have:

$$u_{ijk} = \beta_i^q \bar{q}_i + \beta_i^q \tau_k \Delta q_{ij} + X_{ij} \beta^X + Y_j \beta^Y + \varepsilon_{ij} \quad (5)$$

Where Δq_{ij} is the difference between actual quality and $\bar{q}_i$: $\Delta q_{ji} = q_j - \bar{q}_i$. This last term is what we referred to earlier as the demeaned quality.

Note that the term $\beta_i^q \bar{q}_i$ is constant for each household and so it does not affect each household's choice. Therefore, this term will drop out when we normalize the model to estimate it.

Finally, we would like make explicit that our success index (W), which is the key variable that allows us to separate preferences for quality from the perception of quality, does not appear in equation (4). We will use this variable as information in the estimation of the type probabilities for each household, as we describe later in section 7.4.

6.3. Probability of admittance

The choice set Λ_i represents the schools household i considers for its child. However, households are restricted in their choices to the extent that schools have the right of admittance and might choose not to accept student i . Let AD_{ij} be a binary variable, unobservable to us, that equals one if student i could enter school j and zero otherwise, such that:

$$AD_{ij} = \begin{cases} 1 & \text{if } \varrho_{ij} - \nu_{ij} \geq 0 \\ 0 & \text{if } \varrho_{ij} - \nu_{ij} < 0 \end{cases} \quad (6)$$

where $\varrho_{ij} = M_{ij} \delta$ and M_{ij} contains a series of variables that characterize school j 's selectivity in terms of different criteria (religiosity of parents, whether parents are married or not, an admittance test score, or other criteria) and their relationship to relevant household characteristics (e.g. if parents say they are religious).

Then, assuming that ε_{ij}^{ad} are iid and follow a logistic distribution, the probability of admission is given by:

$$Pr(AD_{ij} = 1) = \frac{\exp(\varrho_{ij})}{1 + \exp(\varrho_{ij})} \quad (7)$$

To have a tractable likelihood calculation when we estimate the model, we assume that ε_{ij}^{ad} 's

are realized before parents make the D_i decision.²³ Moreover, to simplify the solution of the model, we assume that, for any student, there is always at least one school willing to admit him or her.²⁴ In particular,

$$h \in \operatorname{argmax}_{j \in \Lambda_i}(\varrho_{ij}) \Rightarrow AD_{ih} = 1. \quad (8)$$

Finally, let $\mathcal{J}_i$ be the set of schools in Λ_i that admitted student i :

$$\mathcal{J}_i = \{j \in \Lambda_i : AD_{ij} = 1\} \quad (9)$$

Given our assumptions, these are the schools among which household i chooses.

Because AD_{ijt} is unobserved, there is a challenge in separately identifying the parameters governing this selection process from the parameters in the utility function. To do this, we leverage the rich information we have about the admission rules of each school as exclusion restrictions. As we described in section 4, we build measures of the selectivity of schools according to four different criteria and interact them with relevant household characteristics. Thus, these variables appear in the equations for the selection process, but not in the utility function. Similarly, none of the variables in the utility function appear in the selection process.²⁵

6.4. School choice

Let $D_i \in \Lambda_i$ be the school that is chosen by student i among all schools admitting that student, given that student is of information type k . Formally:

$$D_i = \operatorname{argmax}_{j \in \mathcal{J}_i} \{u_{ijk}\} \quad (10)$$

²³This means parents do not apply to schools. Instead, at the end of each period they know their feasible set for the next period and they pick the feasible school that maximizes their expected utility.

²⁴Given that in the data the $\operatorname{argmax}_{j \in \Lambda_i}(\varrho_{ijt})$ is almost certainly a public school (because most public schools do not select students), this assumption is equivalent to assuming that there is one public school that admits the student i without uncertainty. This is actually the way it works in reality.

²⁵The few variables that are shared by both functions are interacted with other variables in the function that determines the probability of admission.

6.4.1. Distributional assumptions

Finally, for the estimation of the model, we will also need to make some distributional assumptions for the error terms in the flow utility function, ε_{ij} 's. We assume $\varepsilon_{ij} \sim N(0, 1)$ and $cov(\varepsilon_{ij}, \varepsilon_{ij'}) = 0 \quad \forall j \neq j', \forall i$.

7. Estimation

7.1. Simulated multinomial probit with endogenous quality

In our model, parents choose schools based on their quality (among other things). However, as discussed earlier, the quality we observe might be endogenous. In particular, we are concerned about reverse causality, due to the fact that changes in enrollment, determined by a higher probability of that school being chosen, might in turn lead to changes in academic quality. To address this, we instrument for school quality in our estimation. As we described in section 4, our instrumental variable (z_j) exploits exogenous variation coming from additional funding available to schools based on enrollment of socio-economically vulnerable students. We have called this the *SEP* instrument.

We estimate the school choice model using simulated maximum likelihood estimation. To begin the description of how we build the likelihood function, notice that our model, econometrically speaking, is similar to a multinomial probit in which the error terms are uncorrelated with each other. There are some key differences, however. First, as we just mentioned, quality is endogenous and we need to instrument for it. Second, we simulate potential acceptance into each specific school, implying we do not have an analytical expression for the likelihood function. This is what makes our estimation procedure *simulated* maximum likelihood (as opposed to maximum likelihood). Finally, we include unobserved heterogeneity in the form of the information/cognitive types. We will now describe how we include each of these three elements in the estimation, using a multinomial probit as a starting point.

7.2. Endogenous quality: using the control function approach

Recall that the perceived quality for household i of type k is defined as (equation (4)):

$$\tilde{q}_{ijk} = \tau_k q_j + (1 - \tau_k) \bar{q}_i$$

and, as shown in equation (5), the indirect utility of each alternative is given by:

$$u_{ijk} = \beta_i^q \bar{q}_i + \beta_i^q \tau_k \Delta q_{ij} + X_{ij} \beta^X + Y_j \beta^Y + \varepsilon_{ij}$$

where Δq_{ij} is the difference between actual quality and $\bar{q}_i$: $\Delta q_{ji} = q_j - \bar{q}_i$.

We need to define a first stage for Δq_{ij} :

$$\Delta q_{ij} = \gamma_0^Z + z_j \gamma^Z + X_{ij} \gamma^X + Y_j \gamma^Y + \eta_{ij} \quad (11)$$

Finally, we need to make some distributional assumptions about the error terms in this reduced form equation for quality (η 's) and how they relate to the error terms in the indirect utility functions. Recall the ε 's are distributed as independent standard normal variables. We also assume:

1. $\eta_{ij} \sim N(0, \sigma_\eta^2)$
2. $cov(\eta_{ij}, \eta_{ij'}) = 0 \quad \forall j \neq j'$
3. $cov(\eta_{ij}, \varepsilon_{ij}) = \rho \sigma_\eta \quad \forall j \quad \forall i$

These assumptions imply we can write:

$$u_{ijk} = \beta_i^q \tau_k \Delta q_{ji} + X_{ij} \beta^X + Y_j \beta^Y + \frac{\rho}{\sigma_\eta} \eta_j + e_{ij} \quad (12)$$

where $\eta_{ij} = \Delta q_{ij} - (\gamma_0 + z_j \gamma^Z + X_{ij} \gamma^X + Y_j \gamma^Y)$, as defined in equation (11). This is precisely the control function approach: conditional on η_{ij} , the demeaned quality is exogenous. Notice we have subtracted, without any consequence, $\beta_i^q \bar{q}_i$ from all utilities for student i . Based on the distributional assumptions we have made, the error term in equation (12) is distributed $e_{ij} \sim N(0, (1 - \rho^2))$.

Finally, it is important to note what can and cannot be identified from the product $\beta_i^q \tau_k$. Recall we have defined $\beta_i^q = \exp(H_i \alpha)$, where H is mother's education, which serves as a proxy for the household's socioeconomic status. Because H is a categorical variable, we can only identify differences with respect to a reference category. Thus, we normalize $\beta_i^q = 1$ for household's where the mother has secondary education. Given this normalization, the values

we estimate for the τ_k 's should be interpreted as the product of the true τ_k 's and the value of β_i^q for household's where the mother has secondary education. In other words, τ_k 's are identified conditional on the β_i^q normalization. Under any other normalization, we would obtain different estimators. This limitation prevents us from interpreting the magnitude of τ_k 's as the relative weight that households of type k give to the true academic quality of school in their perceived quality, as shown in equation (4).

7.3. Simulated choice sets

As discussed in Section 6, our estimation strategy needs to take into account the fact that some schools have selection processes to admit students, which implies that some schools are actually not available for some households. However, we do not observe which schools are truly in a given household's choice set. We address this issue by using information on selectivity of schools and explicitly including the probability of selection when building the likelihood function.

Recall Λ_i is the set of all possible schools that could be *considered* by student i . Similarly, $\mathcal{J}_i$ is the set of schools that student i could *attend*, taking into account the acceptance/rejections from schools in Λ_i . We do not observe $\mathcal{J}_i$, so we simulate it and calculate the likelihood contribution for student i conditional on this simulated choice set. We then build the actual likelihood contribution as the expectation over these possible choice sets. We formally describe this process below.

First, define the indicators of school choice for household i who is of type k , conditional on the actual choice set, as

$$d_{ijk} = 1 \left\{ u_{ijk} > u_{ij'k} \forall j' \neq j \in \mathcal{J}_i \right\}$$

For convenience, let Ω_i be the set of all exogenous variables that enter the utility function: $\Omega_i = \{X_i, H_i, Y_i\}$, where $Y_i = \{Y_j\}_{j \in \mathcal{J}_i}$ and $X_i = \{X_{ij}\}_{j \in \mathcal{J}_i}$. Similarly, $Q_i = \{\Delta q_{ij}\}_{j \in \mathcal{J}_i}$ and $Z_i = \{z_j\}_{j \in \mathcal{J}_i}$.

We can summarize the indirect utilities in equation 12 as:

$$u_{ijk} = V_k(\Omega_i, \Delta q_{ij}) + \frac{\rho}{\sigma_\eta} \eta_{ij} + e_{ij} = V_{ijk} + \frac{\rho}{\sigma_\eta} \eta_{ij} + e_{ij}$$

Thus:

$$P(d_{ijk} = 1 | \Omega_i, Q_i, \mathcal{J}_i) = P\left(e_{ij'} - e_{ij} < V_{ijk} - V_{ij'k} + \frac{\rho}{\sigma_\eta}(\eta_{ij} - \eta_{ij'}) \quad \forall j' \neq j \in \mathcal{J}_i\right) \quad (13)$$

Recall that, as we had defined them earlier, the dummy variables AD_{ij} 's indicate whether student i would be admitted into school j . Similarly, the probability of admission, $P(AD_{ij})$ was defined in equation (7). Given that we assumed the ε 's are distributed independent standard normal, the vector $(e_{i1} - e_{ij}, e_{i2} - e_{ij}, \dots, e_{iJ} - e_{ij})$ is distributed as multivariate normal with mean zero and variance-covariance matrix:

$$(1 - \rho^2) \begin{pmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 2 \end{pmatrix}$$

This allows us to have expressions for the probabilities in equation (13).

Given this structure, the choice probability $P(d_{ijk} = 1 | \Omega_i, Q_i)$, which we need to build the likelihood, is simulated following these steps:

1. For each household i , simulate N sets of possible schools to be chosen $\{\mathcal{J}_i^n\}_{n=1}^N$, by drawing one vector $\{\nu_{ij}\}_{j=1}^J$ for each n .
2. Calculate $P(d_{ijk} = 1 | \Omega_i, Q_i, Z_i, \mathcal{J}_i^n)$ for each n .
3. The simulated choice probability is given by $\tilde{P}(d_{ijk} = 1 | \Omega_i, Q_i) = \frac{1}{N} \sum_{n=1}^N P(d_{ijk} = 1 | \Omega_i, Q_i, \mathcal{J}_i^n)$.

7.4. Simulated Likelihood Function

Given all of the above, the contribution to the likelihood function for student i if he/she is of information type k is given by:

$$\mathcal{L}_{ik} = \prod_{j \in \mathcal{J}_i} \left[\tilde{P}(d_{ijk} = 1 | \Omega_i, Q_i) f(\Delta q_{ij} | \Omega_i, Z_i) \right]^{d_{ij}}$$

Now we make use of the different types. Recall that W , our success index, is the key variable that allows us to identify differences in how households perceive quality. Thus, we define $\pi_k(W_i)$ as the probability that a student with success index W_i is of information type k .²⁶ Thus, the likelihood contribution for student i becomes:

$$\mathcal{L}_i = \sum_k \left\{ \pi_k(W_i) \prod_{j \in \mathcal{J}_i} \left[\tilde{P}(d_{ijk} = 1 | \Omega_i, Q_i) f(\Delta q_{ij} | \Omega_i, Z_i) \right]^{d_{ij}} \right\} \quad (14)$$

where

$$f(\Delta q_{ij} | \Omega_i, Z_j) = \frac{1}{\sigma_\eta} \phi \left(\frac{\Delta q_{ij} - (\gamma_0 + z_j \gamma^Z + X_{ij} \gamma^X + Y_j \gamma^Y)}{\sigma_\eta} \right)$$

For the estimation of the model, we set the number of information types to two (i.e. $K = 2$). When we estimated the model with $K = 3$, we found no differences in the parameters between two of those types. This means that, given the data and under the assumptions of the model, $K = 2$ is sufficient to difference households according to their ability to tell schools apart.

8. Results

8.1. Parameter estimates

Table 7 shows the estimates for the parameters in the model. Although most of the parameters in the utility function do not have a direct interpretation, it is important to note that all of them have the expected signs.²⁷ Of the characteristics that are not related to quality, the estimates suggest the match characteristics - whether the school is in the same neighborhood and whether the school is religious if the family declares to be religious - are of particular importance.

Estimates for the preference and perception heterogeneity are shown as the relevant coefficient - preference or perception - for each of the groups listed. Preferences, for instance, are estimated relative to those of families with *medium education* (those with secondary but not tertiary education). The results show a steep gradient of preferences. The importance of quality, as measured by test scores, on the choices of families with low education appears

²⁶We specify $\pi_k(W_i)$ as a logit function.

²⁷The only exception is “Private without voucher”, that should have positive sign. However, when interpreting these signs, it is important to remember these parameters must be interpreted conditional on other variables, particularly the enrollment before the SEP implementation.

to be lower than for those with medium education. As shown in Table 7, the value of the preference parameter for low education families is half the magnitude of the parameter for medium education families. This should not be interpreted, however, as quality being half as important in decisions. Similarly, the value of this parameter for high education families is 41% higher than for medium education ones. On the other hand, estimates show that our success index is very good at distinguishing families who can and cannot differentiate schools based on their quality (see Table 8). This is not surprising, as this is precisely what the survey intended to measure. Nevertheless, it is reassuring. Based on these estimates, one of the two types has almost no ability to differentiate between schools, while the other one does have some discerning abilities. Recall that, in the model, the perception parameter, τ , measures the relative weight placed on true quality, relative to average quality. Unfortunately, as mentioned before, these τ_k parameters are only identified up to a normalization ($\beta_i^q = 1$ for the *medium education* group), which prevents us from interpreting the magnitudes of the estimated parameters in this way.

[chequear] In term of selection, all the criteria, except for those grouped under “other”, appear to be important. Consistent with the evidence shown earlier in Table 1, the strongest selectivity is the one that comes from skill or knowledge tests.

Table 7: PARAMETER ESTIMATES

	Coef	SE		Coef	SE
Utility: School Characteristics (Y)			Selection		
Private w. voucher	0.076	0.051	Constant	2.698	0.020
Private w/o voucher	-0.127	0.119	Religious	-0.325	0.125
Frac. SEP	-0.087	0.212	Marital	-0.001	0.194
Frac. SEP Squared	-0.534	0.247	Test	-0.492	0.035
Frac. Enrollement 2006	7.205	0.558	Other	-0.060	0.085
Fee	-0.006	0.001	Religious*religious parents	0.010	4.323
Rural	-0.788	0.041	Marital*parents married	0.041	0.096
Frac. Enrollement 2006 Squared	-18.002	2.017	Test*SES=2	0.015	0.042
Frac. SEP * Private	-0.129	0.135	Test*SES=3	-0.024	0.043
			Test*SES=4	1.830	0.148
Utility: School-household interactions (X)			Test*SES=4,5*SES2	0.000	0.624
Religious match	0.298	0.049	Test*SES=4,5*SES3	0.000	0.963
			Test*SES=4,5*SES4	0.000	0.629
			Religious*SES2	0.000	0.625
Utility: Preference heterogeneity			Religious*SES3	0.000	0.852
Low education	0.499	0.083	Religious*SES4	0.000	0.181
Medium education	1.000	.			
High education	1.412	0.350	Distributional parameters		
			rho	-5.229	0.117
Utility: Perception heterogeneity			sigma	-1.362	0.142
Type 1	0.005	0.117			
Type 2	0.256	0.142			
Type probabilities					
Constant	-10.11	0.70			
Success Index	30.10	1.69			

Note: This table presents the point estimates for our model. As discussed, the medium education parameter (utility: preference heterogeneity) is normalized to one. The standard errors are calculated using the approximation of the Hessian given by the mean of the outer product of the scores. This method to estimate standard error is based on asymptotic theory; in particular, it uses the idea that, under general conditions, $\sqrt{N}(\hat{\Omega} - \Omega) \xrightarrow{p} N(0, A^{-1})$, where $A^{-1} = E[H_i(\Omega)]$ is estimated as $\hat{A} = \frac{1}{N} \sum_{i=1}^N s_i(\Omega)s_i(\Omega)'$, such that s_i is the score: $s_i = \frac{\partial l_i}{\partial \Omega}$.

Table 8: PERCEPTION HETEROGENEITY TYPE DISTRIBUTION
(CONDITIONAL ON SUCCESS INDEX)

	Percentage of high type
All	44.5%
Success index $\leq \frac{1}{3}$	15.7%
Success index $> \frac{1}{3}$ & $\leq \frac{2}{3}$	99.7%
Success index $> \frac{1}{3}$	100%

Note: This table presents the probability of being high perception type given the estimated parameters (Table 7), conditional on the value of the Success index.

8.2. Fit

In Table 9 we assess the goodness of fit of our model. In Panel A, we look at the fraction of students attending each of the three types of schools in terms of administration: public, private with voucher, and private without voucher. The model predicts the distribution of administration type very well. This is also true when we look at this distribution by schooling categories.

Panels B and C of Table 9 show the fit in terms of quality of the schools, both in terms of the demeaned quality, which is the variable we use for estimation, and nationally standardized quality, which is not used directly. In Panel B we look at the fit according to schooling categories, whereas we look at categories of the success index in Panel C. As we show, the model fits the overall levels of quality, according to both measures, pretty well. Furthermore, it also captures the gradient in terms of education reasonably well, and the gradient according to success index extremely well.

Table 9: MODEL FIT

By level of education (Panel A)								
Type of administration	All		Low		Medium		High	
	Data	Model	Data	Model	Data	Model	Data	Model
Public	0.4314	0.4229	0.5827	0.5299	0.4591	0.4471	0.2924	0.3202
Private w. voucher	0.5634	0.5717	0.4173	0.4688	0.5394	0.5496	0.6947	0.6692
Private w/o voucher	0.0053	0.0055	0.0000	0.0013	0.0015	0.0033	0.0128	0.0106
By level of education (Panel B)								
School quality	All		Low		Medium		High	
	Data	Model	Data	Model	Data	Model	Data	Model
Standardized (SIMCE)	0.0102	0.0092	-0.1411	-0.1136	-0.0151	-0.0223	0.1468	0.1307
De-meaned (SIMCE)	0.1129	0.1119	0.0865	0.1140	0.1224	0.1152	0.1233	0.1071
By level of the success index (Panel C)								
School quality	All		Low		Medium		High	
	Data	Model	Data	Model	Data	Model	Data	Model
Standardized (SIMCE)	0.0102	0.0092	-0.0392	-0.0347	0.0979	0.0826	0.1300	0.1318
De-meaned (SIMCE)	0.1129	0.1119	0.0746	0.0791	0.1886	0.1732	0.1778	0.1796

Note: This table compares some conditional means of the data with the ones that come from model simulation. To simulate the model we use the estimated parameters (Table 7), with 100 replicates for each data point.

9. Counterfactual Experiments

We now use the estimated model to conduct a series of counterfactual experiments. The goal of these experiments is to assess the relative importance of preferences, perceptions about quality, and restrictions on school choices. First, we quantify the importance of each of these elements in terms of the average quality of the schools that are chosen by families. Second, we look at their importance for choosing public versus private schools. We are particularly interested in differences in both of these results across education groups. In all of our analysis, as we describe below, we pay special attention to the interaction of preferences and the ability to tell schools apart.

In our first counterfactual, we evaluate the role of perceptions about quality alone. We simulate the choices of households as if all of them were of the “high type” (type 2), *i.e.* the type that does differentiate between different quality schools. In Counterfactual 2, we turn our attention to preferences and set the quality preference parameter for all households equal to the preference parameter of the high education group, which is the group that values quality the most, according to our estimates. Then, to analyze the complementarity between perceptions and preferences, in Counterfactual 3 we combine counterfactuals 1 and 2. Finally, in Counterfactual 4 we turn our attention to restrictions and eliminate differences in selectivity, making all households identical to each other with respect to the characteristics that determine selection.²⁸ This fourth counterfactual, however, does not eliminate all differences in access to schools, as many of these differences are already captured in the set of schools available to each household (choice sets). As we will discuss below, the comparison of the results of some of these counterfactuals will allow us to explore the importance of the differences in choice sets as well.

Before analyzing the results from these exercises, it is necessary to add a cautionary note. As we described in section 6, some of the variables in the utility function of our model are equilibrium outcomes, *e.g.*, the school enrollment in 2007. Because we use data on these variables before the implementation of the policy that provides the source of exogenous variation for school quality, we have argued this is not an issue for the estimation of the model. Nevertheless, using these equilibrium outcomes affects the interpretation of the counterfactual

²⁸We fix those characteristics at the mean or mode, depending on the type of variable

exercises. As we have discussed, these equilibrium outcomes, such as school enrollment in 2007, serve as what would otherwise be a school fixed effect that captures long-term school characteristics that families value. To the extent that these characteristics might be correlated to school quality because of some fundamental relation not captured by the model, a long term change in the school quality, or a long term change in how parents value it, could also affect how much households value these other characteristics. In other words, our counterfactual exercises assume that these school fixed effects (proxied by these pre-sample period equilibrium outcomes) are policy-invariant, when they may not be (especially in the long-run). Therefore, our counterfactual results should be interpreted as the effects of equalizing the ability to tell schools apart, for instance, keeping all of these other elements constant. Alternatively, they can be interpreted as the short-term effects of these changes in how parents value and/or understand school quality on school choices.

To take into account that our estimated parameters are random variables, whose standard errors are presented in Table 7, we show the confidence intervals (CIs) at the 95% level, for the differences between the baseline scenario and each of the counterfactual exercises, for all households and for different subgroups. We study how this randomness impacts the precision with which we can estimate the differences between the baseline scenario and the counterfactual exercises, as these differences are our objects of interest. To calculate these CIs, we assume the parameters in the model follow a multivariate normal distribution with mean equal to the estimated parameters and variance equal to the estimated variance-covariance matrix of those estimates, whose diagonal elements correspond to the standard errors reported in Table 7 (i.e. we use the asymptotic distribution of our estimators). In particular, these CI are calculated using a bootstrap procedure, with the following steps: (1) Take 1000 draws for the set of parameters from the described multivariate normal distribution; (2) For each draw, simulate the model under the different counterfactual scenarios and calculate the differences between the baseline scenario and each counterfactual exercise. With these values, we find the respective percentiles in order to calculate the CIs.

9.1. Quality

We begin by looking at the results in terms of the effects of these counterfactuals on academic quality, as measured by the standardized SIMCE math score.

The overall effect of Counterfactual 1, while not huge, is not trivial. Average quality increases by 0.056 standard deviations if all of the households could tell schools apart as well as type 2 households do. Note we are not letting every household to be able to perfectly observe the “true” school quality. Although such a counterfactual would quantify the cost, in terms of average quality, of limitations to the ability to tell schools apart, we believe it is not as informative as the one we carry out. Instead, Counterfactual 1 takes a more conservative stance and sets this ability at the level of the most able households, according to our estimates.

The gains in quality from Counterfactual 1 are not equally distributed across households of different socioeconomic backgrounds, as measured by the education level. For lower education households, average quality increases by only 0.03 standard deviations. This increase is less than half the magnitude of the increase for high education ones, which is 0.071 standard deviations. There are two mechanisms, working on opposite directions, that determine the size of this difference. On the one hand, as the model explicitly captures, there must be a complementarity between preferences for quality and the ability to differentiate schools according to their quality. Thus, an improvement in the ability to tell schools apart would have a larger impact for those households with stronger preferences for quality - the higher education ones, according to our estimates. On the other hand, because the success index is correlated with education, there are more households of type 1 – those with a lower ability to differentiate – in the lower education group. Thus, Counterfactual 1 improves the ability of a larger fraction of low education households, relative to the higher educated ones.²⁹ The relative importance of each of these mechanisms is determined by the parameter values, as well as the different distributions of other relevant characteristics across education groups. Overall, the complementarity effect dominates, making the quality gains larger for high education households.

As the results from Counterfactual 1 show, differences in preferences appear to be very important in determining differences in average quality across households in different education groups. Thus, in Counterfactual 2 we set the value of the preference parameter at the same value – the value of the high education group – for all households. By construction, this

²⁹Our estimates imply that a household having a high value of the success index almost certainly leads to that household being of type 2, as seen in Table 8.

counterfactual has no effect on the average quality of schools chosen by the high education group. Average quality for the low education households under this scenario increases by 0.037 standard deviations. For the medium education group, the gain is slightly larger; it increases by 0.04 standard deviations.

In Counterfactual 3 we combine the two previous exercises: we make all households equal in terms of the ability to tell schools apart (all are type 2) and equal in terms of their preferences for quality (all with the same value as high education). There are several ways to interpret this experiment. One way is to think about it in terms of the effects of eliminating the differences in the ability to tell schools apart after “controlling for” differences in preferences. Thus, we would assess the outcomes of Counterfactual 3 relative to those in Counterfactual 2. In this case, the gains in quality for the low education group are 0.083 standard deviations ($0.0063+0.0769$), 0.085 for the medium education group, and 0.071 for the high education group. These differences, by construction, cannot be explained by differences in preferences. Therefore, the distribution of other characteristics, including the unobserved types, makes the gains larger for the two lower education groups, relative to the high education group. Put in a different way, even after accounting for differences in preferences for quality, households with lower education seemed to be more affected by the restriction in terms of their ability to tell schools apart.

Counterfactual 3 also allows us to look at the changes in average quality relative to the baseline scenario and compare them to the results from Counterfactuals 1 and 2. This exercise provides a measure of the importance of the complementarity between preferences and the ability to tell schools apart. With respect to overall quality, the first two counterfactuals provide gains of 0.056 and 0.0172 standard deviations, respectively. However, when they are combined, average quality increases by 0.096 standard deviations, which is 0.023 (32%) larger than the sum of the changes from counterfactuals 1 and 2. This complementarity is, naturally, more important for lower education groups, which have lower shares of type 2 households and a lower value of the preference parameter. For instance, for low education households, the total increase in average quality from Counterfactual 3 is 0.12 standard deviations, as compared to 0.030 and .037 from counterfactuals 1 and 2, respectively. This comparison highlights the importance of the interaction between preferences and perceptions,

whose relevance extends beyond the mechanical. Because the household’s education and the success index – our measures for socio-economic status and ability to differentiate qualities – are highly correlated, the complementarity amplifies the effects of limitations in the ability to differentiate qualities for low education households.

The magnitude of the effects of Counterfactual 4 are negligible. Overall, there is a 0.004 standard deviation increase in average quality. These gains are concentrated, as expected, among low and medium education households. As it was already suggested by the parameter estimates, this result reflects the fact that selectivity of schools does not constrain households in terms of access to academic quality, *beyond* deeper restrictions that are already captured by the differences in choice sets. These results are in line with the previous literature that uses Chilean data to study the relevance of schools admission prerogatives in students allocations.³⁰

Regarding the importance of choice sets, Counterfactual 3 provides some useful information. Because all households are identical in terms of preferences and ability to tell schools apart under this scenario, and because Counterfactual 4 has already shown selectivity plays no role, the remaining gap in quality between households with different levels of education can only be attributed to two sources: differences in other characteristics affecting choices (i.e., other variables entering the utility function) and differences in mean quality available under different choice sets.

9.2. Pubic or private

We now turn our attention to the effects of these counterfactual exercises on the choice between public and private schools (Table 11). First, we would like to note that, with one particular exception, all of the counterfactuals lead to migration away from public schools. Because we conduct experiments that focus on quality as well as restrictions, our results provide evidence that quality (measured as standardized test scores) has some, potentially minor, role in the choice of private over public schools.

Neither Counterfactual 1 (only perception) nor 2 (only preferences) have any sizable effect on the fraction of households choosing public schools. However, in line with the results on average quality, the combination of both of these counterfactuals (Counterfactual 3) appears

³⁰See, for example, [Arteaga et al. \(2016\)](#).

to have an effect that is relevant in magnitude. Overall, the share of private schools increases by around 3.4 percentage points. This effect is larger than the effect of eliminating differences in selectivity (Counterfactual 4), which does not seem to be relevant at all for the public vs private decision. The size of the effect of this combined counterfactual is larger for low education households (4.1 percentage points) than for high education ones (2.5 percentage points). However, relative to the baseline fraction of household choosing public schools in each group, the changes are of similar magnitudes for both groups (between 7% and 8%).

The results from Counterfactual 4 show no relevant changes in the share of public schools when we eliminate differences in the characteristics valued in the selection process of schools. Among high educated households, the fraction appears to increase slightly, but the magnitude of this change is of no significance. Recall, however, that this counterfactual does not change the choice set for each household. Therefore, the differences in shares of public / private schools that persist after Counterfactual 4 capture deeper causes of socioeconomic segregation, including those caused by budget constraints, that are outside of the scope of what we can analyze with our model.

Table 10: COUNTERFACTUAL SIMULATIONS: IMPACT ON SCHOOL QUALITY

	Baseline	Counterfactual exercises (differences with respect to baseline scenario)			
		1	2	3	4
All households	0.0092	0.056 [0.053; 0.061]	0.017 [0.001; 0.038]	0.097 [0.059; 0.15]	-0.002 [-0.04; 0.005]
By Education					
Low	-0.1136	0.031 [0.027; 0.034]	0.036 [0.018; 0.061]	0.120 [0.079; 0.181]	0.000 [-0.032; 0.007]
Medium	-0.0223	0.061 [0.047; 0.076]	0.020 [-0.012; 0.06]	0.106 [0.049; 0.18]	0.000 [-0.038; 0.009]
High	0.1307	0.071 [0.051; 0.098]	0.000 [0.000; 0.000]	0.071 [0.051; 0.098]	-0.004 [-0.048; 0.001]
By success index					
Low	-0.0347	0.086 [0.081; 0.093]	0.007 [0.000; 0.015]	0.128 [0.088; 0.185]	-0.001 [-0.036; 0.005]
Medium	0.0826	0.000 [0.000; 0.000]	0.037 [0.001; 0.083]	0.038 [0.001; 0.084]	-0.002 [-0.045; 0.005]
High	0.1318	0.000 [0.000; 0.000]	0.038 [0.005; 0.082]	0.038 [0.005; 0.082]	-0.002 [-0.049; 0.006]

Note: This table presents the differences in average school quality – for different groups – between four counterfactual experiments and the baseline scenario. It also presents the average school quality at the baseline scenario (first column), to have a reference point. Counterfactual 1 evaluates what would happen if all households were able to differentiate between different quality schools. Counterfactual 2 sets the quality preference parameter for all households equal to the preference parameter of the high education group: the highest value of the preference parameter according to our estimates. Counterfactual 3 combines counterfactuals 1 and 2. Finally, in counterfactual 4 we turn our attention to restrictions and eliminate differences in selectivity, imposing that households have access to all the schools in their choice set. The confidence intervals (CIs) for these differences between the counterfactual experiments and the baseline scenario, at the 95% level, are in brackets. These CIs are calculated following a bootstrap procedure that assumes the estimated parameters are normally distributed with mean and variance-covariance matrix equal to the point estimates presented in Table 7 and their corresponding variance-covariance matrix.

Table 11: COUNTERFACTUAL SIMULATIONS: FRACTION OF ENROLMENT BY TYPE OF ADMINISTRATION

Counterfactual exercises					
(differences with respect to baseline scenario)					
	Baseline	1	2	3	4
All households					
Public	0.423	-0.019 [-0.021; -0.018]	-0.007 [-0.015; -0.001]	-0.034 [-0.053; -0.02]	0.002 [-0.007; 0.053]
Private w. voucher	0.572	0.018 [0.018; 0.02]	0.007 [0.001; 0.014]	0.033 [0.019; 0.05]	-0.002 [-0.051; 0.006]
Private w/o voucher	0.005	0.001 [0.001; 0.001]	0.000 [0.000; 0.001]	0.001 [0.001; 0.003]	0.000 [-0.001; 0.000]
Low education					
Public	0.583	-0.009 [-0.01; -0.008]	-0.015 [-0.024; -0.007]	-0.041 [-0.061; -0.025]	0.000 [-0.008; 0.046]
Private w. voucher	0.417	0.009 [0.008; 0.01]	0.015 [0.007; 0.024]	0.040 [0.025; 0.06]	0.000 [-0.046; 0.008]
Private w/o voucher	0.000	0.000 [0.000; 0.000]	0.000 [0.000; 0.000]	0.001 [0.000; 0.001]	0.000 [0.000; 0.000]
Medium education					
Public	0.459	-0.022 [-0.028; -0.017]	-0.008 [-0.024; 0.004]	-0.039 [-0.067; -0.018]	0.000 [-0.011; 0.052]
Private w. voucher	0.539	0.021 [0.016; 0.027]	0.008 [-0.004; 0.022]	0.038 [0.017; 0.063]	0.000 [-0.052; 0.011]
Private w/o voucher	0.001	0.001 [0.001; 0.001]	0.000 [0.000; 0.001]	0.002 [0.000; 0.004]	0.000 [-0.001; 0.000]
High education					
Public	0.292	-0.025 [-0.034; -0.019]	0.000 [0.000; 0.000]	-0.025 [-0.034; -0.019]	0.005 [-0.003; 0.059]
Private w. voucher	0.695	0.023 [0.017; 0.031]	0.000 [0.000; 0.000]	0.023 [0.017; 0.031]	-0.004 [-0.056; 0.003]
Private w/o voucher	0.013	0.002 [0.001; 0.003]	0.000 [0.000; 0.000]	0.002 [0.001; 0.003]	0.000 [-0.002; 0.000]

Note: This table presents the differences in the fraction of enrollment by school type of administration – for different groups – between four counterfactual experiments and the baseline scenario. It also presents the fraction of enrollment by school type of administration at the baseline scenario (first column), to have a reference point. Counterfactual 1 evaluates what would happen if all households were able to differentiate between different quality schools. Counterfactual 2 sets the quality preference parameter for all households equal to the preference parameter of the high education group: the highest value of the preference parameter according to our estimates. Counterfactual 3 combines counterfactuals 1 and 2. Finally, in counterfactual 4 we turn our attention to restrictions and eliminate differences in selectivity, imposing that households have access to all the schools in their choice set. The confidence intervals (CIs) for these differences between the counterfactual experiments and the baseline scenario, at the 95% level, are in brackets. These CIs are calculated following a bootstrap procedure that assumes the estimated parameters are normally distributed with mean and variance-covariance matrix equal to the point estimates presented in Table 7 and their corresponding variance-covariance matrix.

10. Conclusions

We have estimated a simple static model of school choice that allowed us to conceptually separate three key elements in this choice: preferences for academic quality, the ability to tell schools apart in terms of their academic qualities, and restrictions faced by households with respect to the sets of schools that are available to them. We then used rich information on how parents choose schools to construct a measure of their ability to tell schools apart, which we combined with several sources of administrative data to estimate the parameters of the model, separately identifying these three key components of the model. This allowed us to perform a series of counterfactual experiments, which we used as means to assess the relative importance of each of these elements in parents' choices and the quality of the schools their children attend.

Our results suggest a very important interaction between preferences and perceptions about quality in determining parent's choices. The parameter estimates indicate a very steep gradient in terms of preferences for school quality according to socio-economic status, as proxied through the household's education. Furthermore, our estimates also suggest a large amount of heterogeneity in the ability to tell schools apart. Because of the natural complementarity between preferences and perceptions in parents' decisions, as well as the fact that our measure of the ability to tell schools apart is highly correlated with socioeconomic status, the interaction of preferences and perceptions about quality plays a key role in explaining how limitations to tell schools apart affect the quality of the schools attended by children of different socioeconomic background. This large role is in stark contrast with that of school selectivity, which has a very minor effect on explaining the quality gap, according to our estimates. However, there is strong suggestive evidence of the importance of deeper constraints, as reflected by households' choice sets, in explaining differences in quality between children of different socioeconomic background.

These results have important implications for market-oriented educational systems. Because parents choices are the key mechanism through which competition among schools might lead to improvements in quality over time, the limitations suggested by our estimates imply relevant frictions in this mechanism. In particular, our results indicate that a very large fraction of the population is constrained in terms of their ability to tell the quality of different

schools apart. Without the possibility that parents actually choose better schools, the ability of markets to reward better quality schools is limited. Furthermore, less educated households are more likely to belong to this more constrained group. Therefore, these limitations go beyond the proper functioning of markets as information aggregation mechanisms, as they also have relevant consequences in terms of equality in access to academic quality.

References

- ABDULKADIROLU, A., P. A. PATHAK, AND A. E. ROTH (2005): “The New York City High School Match,” *American Economic Review*, 95, 364–367.
- AGUIRRE, J. (2018): “How Can Progressive Vouchers Help the Poor Benefit from School Choice? Evidence from the Chilean Voucher System,” Working paper.
- ALDERMAN, H., P. F. ORAZEM, AND E. M. PATERNO (2001): “School Quality, School Cost, and the Public/Private School Choices of Low-Income Households in Pakistan,” *The Journal of Human Resources*, 36, pp. 304–326.
- ALLENDE, C., F. GALLEGOS, AND C. NEILSON (2019): “The equilibrium effects of informed school choice,” Working paper.
- ANGRIST, J., E. BETTINGER, AND M. KREMER (2006): “Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia,” *American Economic Review*, 96, 847–862.
- ARTEAGA, F., V. PAREDES, AND R. PAREDES (2016): “School Segregation in Chile: Residence, co-payment, or preferences?” Working paper.
- ATTANASIO, O. P., C. MEGHIR, AND A. SANTIAGO (2012): “Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA,” *Review of Economic Studies*, 79, 37–66.
- AUGUSTE, S. AND J. P. VALENZUELA (2006): “Is It Just Cream Skimming? School Vouchers in Chile,” Working paper.
- AZMAT, G. AND J. GARCIA-MONTALVO (2012): “The Role of Awareness, Information Gathering and Processing in School Choice,” Working Papers 639, Barcelona Graduate School of Economics.
- BARTIK, T. J. (1991): *Who Benefits from State and Local Economic Development Policies?*, no. wbsle in Books from Upjohn Press, W.E. Upjohn Institute for Employment Research.

- BAST, J. L. AND H. J. WALBERG (2004): “Can parents Choose the Best Schools for their Children?” *Economics of Education Review*, 23, 431 – 440, special Issue In Honor of Lewis C. Solman.
- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 63, 841–90.
- BETTINGER, E. (2011): “Chapter 7 - Educational Vouchers in International Contexts,” in *Handbook of The Economics of Education*, ed. by S. M. Eric A. Hanushek and L. Woessmann, Elsevier, vol. 4 of *Handbook of the Economics of Education*, 551 – 572.
- BLACK, S. E. (1999): “Do Better Schools Matter? Parental Valuation of Elementary Education,” *The Quarterly Journal of Economics*, 114, 577–599.
- BLANCHARD, O. AND L. KATZ (1992): “Regional Evolutions,” *Brookings Papers on Economic Activity*, 23, 1–76.
- BRAVO, D., S. MUKHOPADHYAY, AND P. E. TODD (2010): “Effects of School Reform on Education and Labor Market Performance: Evidence from Chile’s Universal Voucher System,” *Quantitative Economics*, 1, 47–95.
- CARNEIRO, P., K. T. HANSEN, AND J. J. HECKMAN (2003): “2001 Lawrence R. Klein Lecture: Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice,” *International Economic Review*, 44, pp. 361–422.
- CHUMACERO, R. A., D. GÓMEZ, AND R. D. PAREDES (2011): “I Would Walk 500 Miles (if it Paid): Vouchers and School Choice in Chile,” *Economics of Education Review*, 30, 1103–1114.
- CUNHA, F., J. HECKMAN, AND S. NAVARRO (2005): “Separating Uncertainty from Heterogeneity in Life Cycle Earnings,” *Oxford Economic Papers*, 57, 191–261.
- EPPLÉ, D., A. JHA, AND H. SIEG (2018): “The superintendent’s dilemma: Managing school district capacity as parents vote with their feet,” *Quantitative Economics*, 9, 483–520.

- EPPLE, D. AND R. ROMANO (2008): “Educational Vouchers and Cream Skimming,” *International Economic Review*, 49, 1395–1435.
- (2012): “Economic Modeling and Analysis of Educational Vouchers,” *Annual Review of Economics*, 4, 159–183.
- EPPLE, D. AND R. E. ROMANO (1998): “Competition between Private and Public Schools, Vouchers, and Peer-Group Effects,” *The American Economic Review*, 88, pp. 33–62.
- EPPLE, D., R. E. ROMANO, AND M. URQUIOLA (2017): “School Vouchers: A Survey of the Economics Literature,” *Journal of Economic Literature*, 55, 441–492.
- FU, C. AND J. GREGORY (2018): “Estimation of an Equilibrium Model with Externalities: Post-Disaster Neighborhood Rebuilding,” *Econometrica*, forthcoming.
- GAJARDO, F. AND N. GRAU (2019): “Competition among schools and educational quality: Tension between various objectives of educational policy,” *International Journal of Educational Development*, 65, 123–133.
- GALLEGO, F. A. AND A. E. HERNANDO (2008): “On the Determinants and Implications of School Choice: Semi-Structural Simulations for Chile [with Comments],” *Economía*, 9, pp. 197–244.
- GAURI, V. (1999): *School Choice In Chile: Two Decades of Educational Refor*, Pitt Latin American Studies, Reading, Massachusetts: University of Pittsburgh Press.
- HASTINGS, J. S. AND J. M. WEINSTEIN (2008): “Information, School Choice, and Academic Achievement: Evidence from Two Experiments,” *The Quarterly Journal of Economics*, 123, 1373–1414.
- HENIG, J. (1996): “The Local Dynamics of Choice: Ethnic Preferences and Institutional Responses,” in *Who chooses? Who loses? Culture, institutions, and the unequal effects of school choice*, ed. by B. Fuller, R. Elmore, and G. Orfield, New York: Teachers College Press., 93–117.

- HOUSER, D., M. KEANE, AND K. MCCABE (2004): “Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence Using a Bayesian Type Classification Algorithm,” *Econometrica*, 72, 781–822.
- HSIEH, C.-T. AND M. URQUIOLA (2006): “The Effects of Generalized School Choice on Achievement and Stratification: Evidence from Chile’s Voucher Program,” *Journal of Public Economics*, 90, 1477–1503.
- KANE, T. J., S. K. RIEGG, AND D. O. STAIGER (2006): “School Quality, Neighborhoods, and Housing Prices,” *American Law and Economics Review*, 8, 183–212.
- LARRAÑAGA, O. (2004): “Competencia y Participación Privada: La Experiencia Chilena en Educación,” Working Papers wp208, University of Chile, Department of Economics.
- MARK, G. ELACQUA, AND J. BUCKLEY (2006): “School Choice in Chile: Is it Class or the Classroom?” *Journal of Policy Analysis and Management*, 25, 577–601.
- McMILLAN, R. (2004): “Competition, Incentives, and Public School Productivity,” *Journal of Public Economics*, 88, 1871 – 1892.
- MECKES, L. AND R. CARRASCO (2010): “Two decades of SIMCE: an Overview of the National Assessment System in Chile,” *Assessment in Education: Principles, Policy and Practice*, 17, 233–248.
- MIZALA, A. AND P. ROMAGUERA (2000): “School Performance and Choice: The Chilean Experience,” *The Journal of Human Resources*, 35, pp. 392–417.
- NAVARRO, S. (2010): *Control Functions*, London: Palgrave Macmillan UK, 20–28.
- (2011): “Using Observed Choices to Infer Agent’s Information: Reconsidering the Importance of Borrowing Constraints, Uncertainty and Preferences in College Attendance,” University of Western Ontario, CIBC Centre for Human Capital and Productivity Working Papers 20118, University of Western Ontario, CIBC Centre for Human Capital and Productivity.
- NEILSON, C. (2013): “Targeted Vouchers, Competition among Schools, and the Academic Achievement of Poor Students,” Working paper.

- SÁNCHEZ, C. (2018): “Skipping you Exam? The Unexpected Response to a Targeted Voucher Policy,” Working paper.
- SCHNEIDER, M., P. TESKE, M. MARSHALL, AND C. ROCH (1998): “Shopping for Schools: In the Land of the Blind, The One-Eyed Parent May be Enough,” *American Journal of Political Science*, 42, pp. 769–793.
- SIMON, H. A. (1986): “Rationality in Psychology and Economics,” in *In Rational Choice: The Contrast Between Economics and Psychology*, ed. by R. M. Hogarth and M. W. Reader, Chicago: University of Chicago Press, 25–40.
- VALENZUELA, J. P., C. BELLEI, AND D. D. L. RÍOS (2014): “Socioeconomic School Segregation in a Market-Oriented Educational System. The Case of Chile,” *Journal of Education Policy*, 29, 217–241.
- WOLPIN, K. I. AND P. E. TODD (2006): “Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility,” *American Economic Review*, 96, 1384–1417.