When parents skip schools: evidence of suboptimal choices in the Chilean school admission system

Abstract

This research uses the case of Chile, a well-known experiment of country-wide school choice and voucher policies, to assess potential drawbacks in the application process if parents face a vast school supply. I use data from the new National Admissions System, which allows parents to apply to any public-funded school in the country. The enormous amount of information that applicants face when comparing alternatives could conduct them to omit some more desirable schools, leading to suboptimal individual and social results. By comparing utility levels between stated and potential preferred schools (i.e., the school that maximizes the utility level), it is possible to evaluate if parents are contrasting optimally or are missing relevant alternatives due to the high number of alternatives where to compare from. The main findings suggest that more than 80% of parents omit more desirable alternatives, from which 20 pp. correspond to those who skipped a better alternative closer than the stated preferred school. On average, families would be willing to apply to more distant schools if they receive compensation in other educational attributes, such as a lower proportion of low-income students in those schools and better academic quality.

1.0 Introduction

The School Choice system proposed by Friedman (1962) encourages parents to apply to schools according to the school characteristics that they value the most. Previous research (Hofflinger, 2020; Burgess et al., 2015) has indicated that some of those attributes are teaching quality, the educational curriculum (religious or single-gender education), the access costs (tuition or monthly stipend), among others. However, in an admission system where there are no geographical boundaries or other limits that reduce the number of schools that parents must choose from, the overwhelming amount of information that parents must face may lead them to omit more desirable alternatives.

Skipping schools may lead to individual and social undesired outcomes if those omitted schools are better than those included in the application list by the parents. Since the previous literature has suggested heterogeneous preferences across socioeconomic groups (Burgess et al., 2015; Hastings et al., 2009), I argue that the type and number of skipped schools may also be different across these groups. Parents from different income levels may consider different types of schools as part of their choice sets, implying that a relevant omitted school for some parents may not be considered by another type of parents. These indications would pose a problem in educational markets where the value of school attributes is heterogeneous in space, shaping access to a different school supply depending on the place of residence and socioeconomic level. Therefore, it is relevant to study the effects of the heterogeneity in the parent's enrollment preferences, the attributes that weigh those decisions, the omission of more desirable schools, and the type of schools that they are skipping.

As a contribution to the literature, this research studies how the place of residence and the omission of relevant alternatives can lead to a suboptimal selection of schools. I estimate the parents' utility function for school characteristics and use the utility maximization criteria to identify those schools that would be more desirable than the one reported as most preferred in the submitted list. In school admission systems where parents face many alternatives, some families probably skip

schools that would match better with their school preferences. By contrasting the attributes of the schools reported as most preferred in the admission process and the ones that maximize the parent's utility level, it is possible to measure the consequences of skipping those relevant alternatives.

To create the counterfactual scenario with complete information, this research estimates a utility function to predict the utility level that parents would receive in each alternative according to the observed school attributes. To analyze the school characteristics that families find more relevant, the investigation will follow the literature of the "Random Utility Model" (McFadden, 1977). The empirical approach will be through a Conditional Logit of the first preference submitted by those parents that applied for grades between pre-kindergarten and first grade in the city of Santiago in 2019. To study the hypothesis of heterogeneous preferences for school attributes, the model is also including interactions according to the family's socioeconomic level.

The main findings confirm what is suggested by the previous literature, where attributes of schools such as home-school distance, monetary payments, and high low-income student's composition are undesirable by households. On the other hand, high levels of academic quality and religious teaching are positively valued by applicants. By comparing the predicted utility level of each parent for each school, I find that 80% of applicants skip a school with a higher utility level. For those parents whose first preference in the application list was at most 5 kilometers away, the mean number of alternatives equal or closer as the preferred school was 35 for parents who skipped a relevant better alternative, and 16 for those who chose optimally. These findings imply that for those parents living in sectors with a wide school supply the costs of acquiring more information represent for making optimal decisions may be higher than in sectors with a smaller supply.

2.0 Literature review

Information is a key factor in making decisions, and incomplete information may lead to suboptimal choices if relevant data is omitted when comparing alternatives. In a School Choice system, where parents play an important role, preferences and information conditionate the relevant school supply that parents consider. The previous idea suggests that both the inability to handle large amounts of information and preferences for school attributes can lead to suboptimal individual and social results.

The literature has pointed out certain factors that condition how households compare and list schools, evidencing problems with the School Choice system. Among the most pointed criticisms is the socioeconomic segregation that derives from the heterogeneous preferences for school attributes (Billingham & Hunt, 2016; Oosterbeek et al., 2021), and the residential location (Owens, 2020; Taylor & Gorard, 2001). Although it is possible to avoid the problem of selection-power through admissions systems where schools do not participate in the selection process, heterogeneous preferences and the spatial distribution of households and schools may continue to influence the results.

Regarding household preferences for school attributes, it is argued that dissimilar preferences between social groups can increase the school segregation levels (Oosterbeek et al., 2021; Weiher & Tedin, 2002). It is pointed out that families of distant socioeconomic status would self-select in

different types of schools according to their preferences for school attributes, which would differ in intensity among these groups.

Some relevant attributes correspond to home-school distance, where the literature has indicated that lower-income parents value more the distance between home and school (Burgess et al., 2015; Chumacero et al., 2011). It is suggested that this barrier may be related to direct and indirect costs associated with commuting, such as the cost of transportation, travel time, and security. In the case of more vulnerable families, as they do not have safe means of transportation, they opt for establishments close to their locations (Warrington, 2005). Other studies have also found strong preferences for schools with high proportions of students with similar characteristics to their own (Oosterbeek et al., 2021; Bonal & Bellei, 2018). This would suggest that households would be influenced to select based on their beliefs and desires to surround themselves with other similar individuals.

While heterogeneous preferences can lead to socioeconomic segregation, it is possible to argue that the omission of schools with equal or better characteristics may lead to other social and individual suboptimal results. Research has sought to quantify the negative effects of the incapacity of households to compare the whole school supply, through experiments that allow observing and contrasting social and individual results when choosing in a more informed way. Some studies that used the Randomized Controlled Trial approach suggest that more informed families change schools to choose, preferring those high-performing schools (Valant & Weixler, 2020), more distant from their residential location and with a higher price (Neilson et al., 2019). Other research has suggested that giving more information about the academic quality, funding opportunities, and earning premiums of college alternatives may not have a substantial effect on the modification of the enrollment choices (Bonilla-Mejía, Bottan & Ham, 2019).

The literature has suggested possible factors for not being able to compare the whole school supply. An investigation conducted in Ghana suggests that the costs of getting information are high and lead to losses in well-being, with a possible reduction of 75% in the distance to travel one of these (Ajayi & Sidibe, 2020). The study authors point out, through counterfactual analysis, that 58.7% of the loss of well-being can be attributed to the lack of information from the households due to the high cost of gathering more information from more alternatives. Another investigation argues that part of the effect may be due to the peer effect (Dustan, 2013; Elacqua & Fabrega, 2004). It is suggested that a household's social networks can provide information on certain alternatives and not on others, leading to choosing schools that are not necessarily optimal according to their preferences.

Both preferences for school attributes and the household's level of information on the school supply play a relevant role when choosing schools. Understanding the individual and aggregate consequences of these factors in a School Choice system could help improve it. This research seeks to contribute to the study of preferences and the delivery of information by studying the case of the new Chilean school admission system. The investigation will focus on measuring the consequences of heterogeneous preferences and lack of information in the school attributes that each household would observe by comparing the real information with a counterfactual scenario.

3.0 The Chilean educational context

Over the last 40 years, the Chilean educational system underwent two profound reforms focused on giving households greater choice power over their child's education (School Choice system). The first reform, applied in the early 1980s, introduced the model through a voucher system in subsidized public founded schools (Gauri, 1999). This idea proposed to hand over decision-making power to households through a per-pupil attendance subsidy to schools (Friedman M., 1962). Families would apply to those schools that best fit their preferences, where a fundamental assumption was that households were utility maximizers. The goal of this model was for schools to compete for students through educational plans that were attractive to households, such as good teaching and learning, the educational curriculum, and low monetary costs. This would incentivize lower-performing schools to improve the education quality or close.

However, the literature has revealed multiple problems related to the model application. For example, the introduction of vouchers in the educational system may have had ambiguous effects, since there is insufficient evidence to affirm that it led to an improvement in the quality of public education (McEwan & Carnoy, 2000; Bronfman, 2007; McEwan, Urquiola, & Vegas, 2008). Another relevant problem was the growth of the selection power of schools with good academic performance, due to the excess demand for them. Between 1980 and 2015 schools were able to select applicants (to some extent). Through admissions tests, interviews, requests for evidence of household income, or other signaling strategies, schools avoided students that could lower the school's academic performance. This power allowed them to select those students who could perform better in academic tests, which was strongly related to the student's socioeconomic level. Because of this, the socioeconomic distribution of students between schools evolved to one in which most of the high-income students attended those schools with better teaching quality, while more vulnerable students were enrolled in those with deficient quality.

The second reform was implemented four decades later (2015). This reform included the creation of a new School Admission System ("Sistema de Admisión Escolar" or SAE), which aimed to eliminate the selection power of public schools. This new system follows the Deferred Acceptance algorithm (Gale, D. & L. S. Shapley, 1962), which allowed maximizing the overall utility level while accounting for household preferences. By delegating the student selection task to a computerized system with random components, selection and discrimination problems were avoided while maintaining the school choice system. Currently, families no longer need to apply in person to schools, but rather the process of selection and admission of establishments is done through a website that allows households to create a list of preferred schools (ranked). It should be mentioned that the selection of alternatives does not have restrictions in terms of application or distance to travel limits. Also, the Ministry of Education established categories of preferences during the admission process, such as giving preference to a certain number of low-income students. Together with the number of vacancies per school, the state's priority ranking, and the household's list of preferred schools determine the distribution of students among schools.

Despite the specifications of the new admission system, recent evidence suggests that there was no significant change in the socioeconomic composition of schools within the first 5 years of operation. This situation could be due to high co-payment in schools with higher teaching quality, being a barrier for low-income families (Sillard, Garay, & Troncoso, 2018). Another article states that this

result is due to the minimum proportion (15%) of low-income students per school (Eyzaguirre, Hernando, Razmilic & Blanco, 2019). The researchers suggest that expanding it to 20% in the best-performing schools could reduce access inequality to better education quality. They also point out that the lack of information on the admission mechanism and the available alternatives can condition the speed at which these changes occur in socioeconomic compositions.

Although these researchers suggest multiple reasons that could explain the observed results, they also indicate that part of the effect may be due to the household's preferences for school characteristics, the difficulty of comparing all alternatives, and school's capacity restrictions. They suggest that the differences in information and close alternatives may lead to an unfavorable equilibrium state (in terms of utility maximization) for low-income families. Neilson, Allende & Gallego (2019) may corroborate this idea by pointing out that families from the low-socioeconomic background who participated in an intervention (where they were helped to choose in a more informed way) were substantially benefited. On average, these households could have access to better teaching quality schools, with these results being slightly reduced due to the school's capacity constraints.

4.0 Data

To study parents' preferences for school characteristics, this research uses public and administrative data from the application process in the new admission system (SAE for now on) provided by the national Ministry of Education. Family's and schools' geographical locations are available, as well as attributes of students and alternatives. To address the possible socioeconomic heterogeneity of parents' preferences, this research uses two proxies to separate families and students into different groups. I use a block-level Socioeconomic Index and a dummy variable of whether the student is from a low-income family or not to sort families in different GSE groups.

4.1 Information of parents and students

This research focuses on those parents from 2019 who applied to grades from pre-kindergarten to 1st grade in the city of Santiago[1], Chile. This supposes a group of 80,772 students, which represents 16.7% of the total applicants in 2019 and 56.6% of those residing in the city of Santiago. It is necessary to take into consideration that the new SAE considers only public-funded schools. Therefore, richer families that can apply to private schools are not included in the dataset.

The databases published by the Ministry of Education contain information about the students, as well as the list of schools nominated by each parent, sorted by preference. The available data contain information about the family's location, the grade to which it applies, and a binary variable indicating whether the student is a priority student or not[2]. This label is assigned according to the family's socioeconomic level and is used as one of the approximations of its socioeconomic level. With the parent's location, it is possible to compute the commune of residence and if the family

[1] Comunes that are included in Santiago: Santiago, Cerrillos, Cerro Navia, Conchalí, El Bosque, Estación Central, Huechuraba, Independencia, La Cisterna, La Florida, La Granja, La Pintana, La Reina, Las Condes, Lo Barnechea, Lo Espejo, Lo Prado, Macul, Maipú, Ñuñoa, Pedro Aguirre Cerda, Peñalolén, Providencia, Pudahuel, Quilicura, Quinta Normal, Recoleta, Renca, San Joaquín, San Miguel, San Ramon, Vitacura, Puente Alto, San Bernardo and Padre Hurtado.

[2] Among the categories to be selected as a priority student are: (i) belonging to the 33.3% of the most vulnerable population of the Chilean Social Household Registry, (ii) being in section A of the National Health Fund or (iii) belonging to the Chile Solidario Social Protection System, Ethical Family Income or the Security and Opportunities System. In case of not fulfilling any of the previous cases, it is considered household income, level of education of the parents and level of rurality and poverty of the area where the student resides.

lives in a rural or urban sector. Since the perception of school characteristics may vary between individuals from urban and rural sectors (e.g., distance perception) and grades (e.g., older students may prefer better school quality to travel distance compared to younger students who need to be accompanied by their parents), the study focus on those students from pre-kindergarten to 1st grade who were living in the urban area of Santiago by the time they participated in the admission system. That means that, from the 483,070 students that applied to a school in 2019, 80,772 fit the described criteria.

4.2 Schools

Since this research only considers families located in the urban sector of Santiago, the schools that represent the set of available alternatives are limited to a similar territorial unit. This means, among other things, that parents applying to schools in other regions or distant from the city are not considered in this study. However, it is possible to argue that parents close to the urban limits can apply to nearby schools outside of the said territorial unit (e.g., rural areas or suburbs). Consequently, and since on average students do not usually travel more than 2 kilometers away from home, schools that are within a maximum of 5 kilometers from the urban border are also included. Annex 8.1 plots the previous idea, where a buffer around the city of Santiago includes all the schools that are considered within the set of available alternatives. Given this, and since the research only considers those schools that dictate the grades from pre-kindergarten to 1st grade, the set of available alternatives in the school market corresponds to 1,211 establishments.

Given the previous literature, each school includes information about several school characteristics. Available data include the SIMCE math test score (as a proxy for academic quality), the monthly stipend cost, the proportion of priority students (as a continuous proxy for the school's low-income composition), the mean number of students per classroom, a binary variable of whether the school has single-gender education and another if it has religious education. Finally, the location of each alternative is included. This information allows calculating the distance between each parent and available school.

4.3 Center of Territorial Intelligence

Instead of using the variable of whether the student is a priority student or not as a label of the family's socioeconomic level, the percentile of the household's neighborhood from 2012 is used. The data comes from an index designed by the Center of Territorial Intelligence, an institution part of the Universidad Adolfo Ibáñez. For each block, this indicator considers the educational level of the head of the family, the house quality, and the acquired goods for its preparation.

Since the parent's location given by the Ministry of Education has a random error between 50 to 300 meters from the actual location, the percentile of those blocks whose centroid is within a radius of 300 meters from the reported location is weighted. This allows an approximation to the family socioeconomic through spatial information.

4.4 Data analysis

Anex 8.2 shows the characteristics and definitions of the variables to be used in this research. Regarding the approximations of the household's socioeconomic level, a proportion of 35.5% priority students are observed. On the other hand, the average socioeconomic percentile of the

household's neighborhood (within a radius of 300 meters) is 58.2%. It should be noted that according to the neighborhood's socioeconomic percentile (external source to the one provided in the admission process), 16 applicants are lost due to a lack of information.

Regarding the characteristics of the available offer, a statistical summary with the variables to be considered as relevant by the households it is also made. It is seen that 29.1% of schools have an educational curriculum with religious education, while only 4% have differential education. Although more than 75% of the available schools do not have co-payment (since a part of these is municipal schools and another are private subsidized schools without co-payment), the average of $11,075 Chilean pesos observed is influenced by a significant co-payment amount in the upper part of the distribution. For the case of the proportion of priority students within the schools, the distribution tends to be a normal one with a mean of 46.1% of priority students and a standard error of 18.2 percentage points. Finally, the distribution of the SIMCE math test score has a slight positive bias to the right, concentrating the bulk of schools in values below the mean (252 points).

Figure 01: Spatial distribution of school attributes and household's socioeconomic level

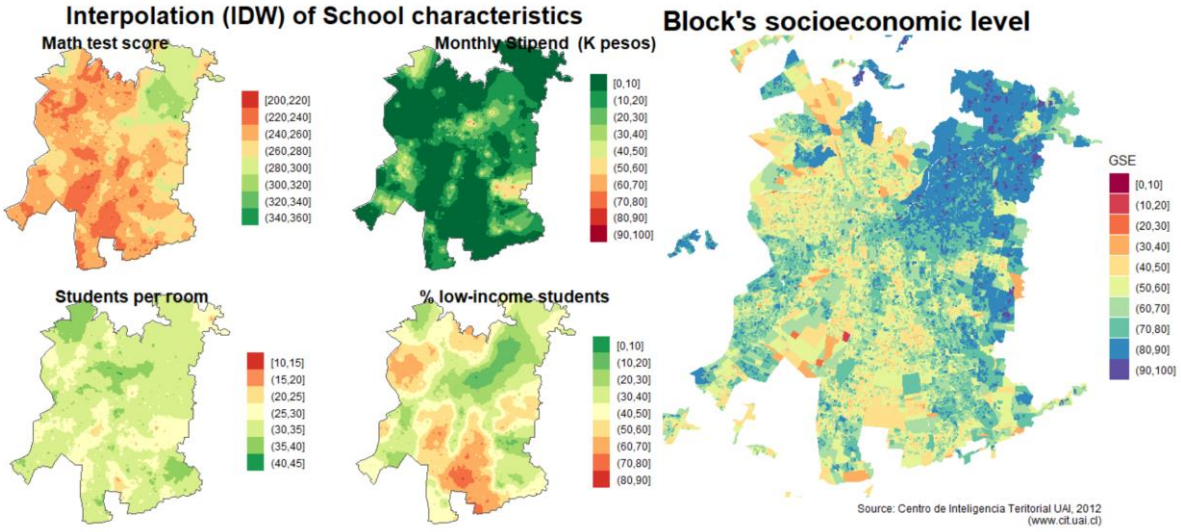


Figure 01 show the spatial distribution of the main school attributes, as also the block's GSE distribution. The high spatial correlation between the school's academic quality and the neighborhood's socioeconomic level denotes some consequences of urban socioeconomic segregation. If parents are discouraged from considering further alternatives, the place of residence and the characteristics of the close school supply may conditionate the academic quality that they can choose from.

5.0 Methodology

Following the Random Utility Model literature (McFadden, 1977), this research seeks to measure to which degree parents can choose optimally among a wide school supply. By estimating parents' preferences for educational attributes, it is possible to construct a utility function that allows sorting schools according to the predicted utility level that parents would receive from each one of them. By finding a school that would bring parents a higher utility than the one selected as most preferred

in the application list, it is possible to suggest that an enormous amount of information could be detrimental to choosing optimally.

The empirical approach consists of the estimation of a utility function, where a Conditional Logit of the parent's first preference indicated in the SAE application list is contemplated. The model includes those school attributes indicated by the literature as most relevant by parents, as also interactions of each attribute with the family's socioeconomic level. This will allow us to test if heterogeneous preferences across socioeconomic groups may lead parents to consider a different school supply as also skip different types of schools.

5.1 Framework

The School Choice literature has modeled parent's preferences using the Random Utility Model approach (McFadden, 1977), which is often used in mutually exclusive alternatives election problems. This model assumes that household "i" choose a school from a set "S" of alternatives, from which he obtains a certain utility level (such that $U_{i,j} \geq U_{i,k} \forall j,k \in S\ (j \neq k)$). Said level of utility of individual "i" over a school "j" can be linearly decomposed into:

$$U_{i,j} = \delta' D_{i,j} + \mathrm{X}_j' \beta + \varepsilon_{i,j}$$

Where X_j is a vector that contains the selected school "j" specific attributes, while $D_{i,j}$ represent the home-distance attribute between family "i" and school "j", β and D represents the vectors of parameters to be estimated and $\varepsilon_{i,j}$ is a stochastic error not observed by the researcher. Those parameters represent the marginal utility of an increase in each of the school attributes. School choice studies often model these functions using classic discrete choice models, such as the Conditional Logit (CL) or the Rank-ordered Logit (ROL). In the CL, the probability that a parent "i" chooses school "j" can be expressed as:

$$\Pr(\mathrm{Y_i} = \mathrm{j}) = \frac{\exp\left(\delta' D_{i,j} + \mathrm{X}_j' \beta\right)}{\sum_{j=1}^{J} \exp\left(\delta' D_{i,j} + \mathrm{X}_j' \beta\right)}$$

Since $\varepsilon_{i,j}$ is unknown, the result is expressed in terms of probability under the assumption that $\varepsilon_{i,j}$ distributes IID. This assumption leads to the imposition of the independence of irrelevant alternatives property (IIA), which establishes that the ratio of choice probabilities for two alternatives is not affected by the addition of a third one:

$$\frac{P(A_z|S')}{P\left(A_y|S'\right)} = \frac{P(A_z|J)}{P\left(A_y|J\right)} \quad z, y \in S'(z \neq y), \forall S' \subseteq J$$

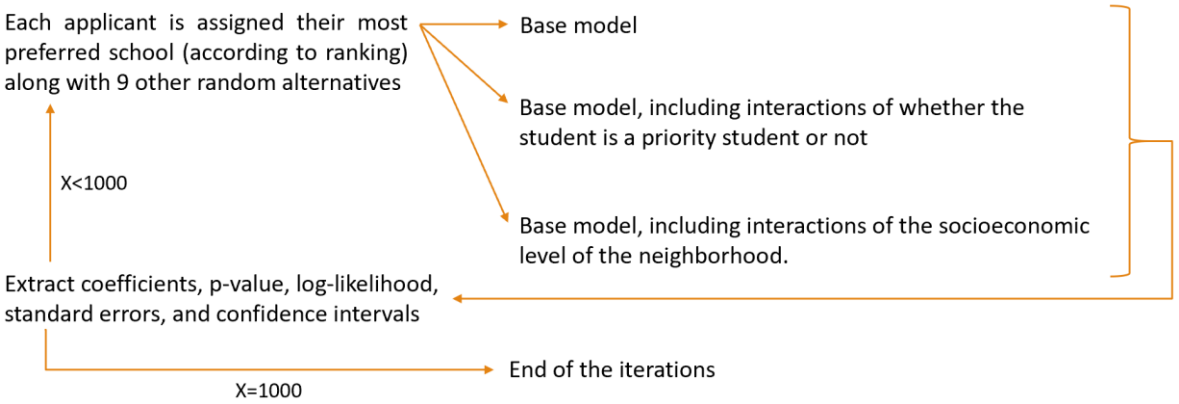
Where A_j corresponds to the alternative "j" and S' is a set of alternatives within all Schools J. Consequently, a change in the probability of an alternative will lead to proportionally equal variations in the relative choice probabilities for all the other alternatives.

5.2 Empirical strategy

Since parents face an election problem, this study uses a Conditional Logit to estimate preferences for school characteristics of families applying for grades between pre-kindergarten and 1st grade in the city of Santiago. Given the nature of the application process, where parents do not have a

minimum or a maximum number of schools to apply to, the research considers only the first preference indicated by them to study the entire target population. This supposes a universe of 80,772 applications and 1,211 available schools.

Figure 03: Bootstrapping estimation process (1000 iterations)



Given the characteristics of the problem, where parents must choose among a wide range of alternatives, some computational barriers make it difficult to directly estimate the utility function. To sort this problem, I follow López & Greenlee (2016) approach and estimate the models by using bootstrapping of a random sampling of alternatives to observe the empirical distribution of the coefficients. Figure 03 describes the bootstrapping process. For each iteration, out of a total of 1,000, each parent compared the most preferred alternative indicated in their application list with the other 9 schools randomly assigned. These are replaced in each iteration by the 9 random alternatives so that everyone compares the information of many schools with the one selected as favorite. For each school attribute, it is possible to observe the distribution of the 1,000 iterated coefficients. By observing the existence of convergence of these distributions, it is possible to assign the converged values like the ones that will be included in the utility function. For the comparability of the bootstrapping process across specifications of the Conditional Logit, the same subsamples of random alternatives are used in all the specifications of the utility function to be estimated. This assumes that an individual "i" faces, in the same iteration, the same subsample in the base model and the ones that include interaction terms of the family's socioeconomic level.

The research includes an after-estimation analysis that uses the estimated utility function to test if parents are omitting more desirable alternatives. The base specification of the model will only include the school attributes as part of the utility function, acting as a benchmark for the following specifications that test parents' heterogeneous preferences across socioeconomic groups. The literature has recognized distance, academic quality, monetary cost, and socioeconomic/racial composition as the aspects most valued by parents when choosing a school (Hofflinger, 2020; Burgess et al., 2015; Hastings et al., 2009). Consequently, the base model includes home-school distance (kilometers), the school's average SIMCE math test score (as an approx. for learning and teaching quality), monthly stipend costs (in thousands of Chilean pesos), percentage of priority students (a proxy for the proportion of low-income students at school), a binary variable of whether the school has religious education and another of whether it has single-gender education. Annex

8.2 shows the definition of each variable to be used. The utility function of parent "i" on school "j" can be expressed as:

$$U_{i,j} = \delta' D_{i,j} + \mathrm{X}_j' \beta + \varepsilon_{i,j}$$

The previous expression corresponds to the base model to be estimated, where the vector X_j contains the previously mentioned attributes for school "j", while $D_{i,j}$ represent the home-school distance between parent "i" and school "j". Recognizing the previous findings of the literature that suggest heterogeneous preferences across socioeconomic groups, I estimate two alternative specifications of the base model that include socioeconomic interactions. I add to the base model socioeconomic interactions to the school attributes, these being the block's socioeconomic percentile and the category of whether the student is a priority student or not.

5.3 After estimations

Our first approach to measure parents' inefficient choosing will consider the omission of clone schools that are closer to the location of residence. By using the representative coefficients of the utility function, it is possible to define a clone school as an alternative that has equal or higher values of desirable school attributes and equal or lower values of undesirable characteristics. If distance represents an undesirable school attribute, finding clone schools that are closer to the parent's location than the one selected as most preferred in the application list will necessarily bring them a higher utility level. Therefore, by finding the proportion of parents that skip closer more desirable clone schools it is possible to measure a proxy for the inefficiency of the parent's choice capability.

A second proxy will consider the trade-off of school attributes that parents face when comparing many alternatives. It is possible to use the previously estimated utility function to assign, for each parent, the utility level that they would receive in each school according to the observed school attributes. This will allow us to find a school that maximizes their utility level that not necessarily is close to the parent's location. By contrasting the information of this school with the one indicated as the most preferred in the submitted application list, it is possible to detect those parents whose optimal school is not the one included as most preferred in that list. If a significant number of parents in this situation are found, this may be evidence that parents do not compare all the available alternatives optimally, leading to skipping some more desirable schools.

6.0 Results

6.1 Parents' preferences for school characteristics

First, I proceed to describe the results of the bootstrapping process. The 1000 iterations of the three specifications of the utility function show a normal distribution for each of the coefficients. Given this, the mean coefficient of each school attribute is used as the representative value for the construction of the utility functions. Figure 04 includes the mean coefficients, mean standard deviation, mean p-value, and mean log-likelihood of the bootstrapped conditional logits. Since the values of the school attributes are represented by the rough coefficients of the logistic regression and they are in utility terms, I proceed to describe their signs.

For the base model in column 1, I observe that parents are averse to high values of home-school distance, monthly stipend, the proportion of low-income students (proportion of priority students), and if the school limits for one-gender enrollment. On the other hand, parents will prefer those

alternatives with high academic quality (math test score), number of students per room, and if the school imparts religious teaching. Most of these results are according to what was suggested by the previous literature and what it was expected, except the positive preferences for more students per room. While previous research indicates that more crowded classrooms have negative consequences on the teaching and learning capability (Ijaiya, Y., 1999; Khan, P., & Iqbal, M., 2012), these positive preferences for this school attribute may suggest a contradiction between the interests of the parents and the welfare of the student.

Figure 04: Parent's utility function

Parents' Preferences For School Characteristics

Utility	Base	Priority Students		GSE			
		YES (Base)	NO (Interaction)	0%-50% (Base)	50%-70% (Interaction)	70%-100% (Interaction)	$ß_{50\%-70\%}=ß_{70\%-100\%}$ Test's P-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Distance	-0.789***	-0.886***	0.147***	-0.947***	0.171***	0.384***	0.000
(Kilometers)	(0.004)	(0.008)	(0.009)	(0.010)	(0.012)	(0.013)	
Standardized Math Score	0.014***	0.011***	0.004***	0.014***	0	0.003**	0.002
(Points)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
Monthly Stipend	-0.005***	-0.007***	0	-0.004***	-0.003**	-0.004**	0.317
(Thousands Of Pesos)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
% Priority Students	-0.028***	-0.009***	-0.032***	-0.016***	-0.018***	-0.041***	0.000
(Percentage Points)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	
Students Per Room	0.06***	0.07***	-0.016***	0.071***	-0.009**	-0.034***	0.000
(Students)	(0.001)	(0.003)	(0.003)	(0.003)	(0.004)	(0.005)	
Religious School	0.137***	0.227***	-0.123***	0.269***	-0.126**	-0.425***	0.000
(1 = Religious)	(0.016)	(0.028)	(0.035)	(0.035)	(0.040)	(0.057)	
Single Gender School	-0.613***	-0.549***	-0.092	-0.378***	-0.283**	-0.308**	0.588
(1 = Single Gender)	(0.035)	(0.066)	(0.078)	(0.076)	(0.088)	(0.118)	
Obs. Per Category	80772	28694	52078	19682	49875	11199	
Observations	80772	80772		80756			
Mean Log-likelihood	-34002.07	-33072.62		-33082.67			
Alternatives	10	10		10			
Iterations	1000	1000		1000			

Mean Coefficients of Bootstraped Conditional Logit With Random Sampling of Alternatives
Mean standar errors in parentesis
Significance: *** (99%), ** (95%), * (90%)

By comparing the results of the models that address possible heterogeneous preferences across GSE groups (columns 2 to 6), it is possible to observe that both specifications are consistent with the main findings in the base model and between these two proxies of the family's GSE level. The effect of the home-school distance and proportion of low-income students over the parent's utility is according to what is suggested by Burgess, S. et al (2015). I find that parents from higher GSE are more willing to travel a longer distance and are more averse to higher concentration from low-income students in schools, compared to parents from lower GSE. Also, parents from the highest 30% of the socioeconomic distribution have stronger preferences for academic quality, compared to those families from the lower 70%. But contrary to what was found in previous national research

(Hofflinger, A. et al, 2020), the heterogeneous preferences for religious teaching suggest that parents from higher GSE prefer less intensely this school attribute. Those parents from the highest 30% of the GSE distribution prefer schools that do not include religious teaching in their curriculum.

This duality where parents from higher socioeconomic groups are more willing to travel longer distances and value more academic quality may give them more flexibility to choose among a wider school offer. On the other hand, parents from lower GSE levels are limited by their budget constraints and discouraged by their preferences. These differences between socioeconomic groups may lead them to consider different types of schools, leading to suboptimal results where families from different GSE face different alternatives with different school attributes.

Figure 05: Willingness to travel

GSE	Math test score	Monthly stipend	% low-income students	Students per classroom	Religious teaching	Single-gender teaching
0%-50%	14.81	-4.10	-16.59	74.63	283.73	-399.22
50%-70%	16.81	-8.73	-43.29	79.04	183.41	-852.38
70%-100%	29.15	-14-24	-100.71	65.90	-278.46	-1,220.23

By calculating the ratio between the coefficients of the school attributes over the home-school distance coefficient it is possible to observe the willingness to travel. Figure 05 show the willingness to travel for parents of each socioeconomic group. The previous table shows how many extra meters does a parent of each socioeconomic level has to travel to increase in one unit each of the other school attributes (ceteris paribus). I observe that parents from the highest 30% of the socioeconomic distribution are more willing to travel longer distances to choose a school with one extra point in the math test score of the school, compared to families from lower socioeconomic groups. But this same group requires a higher compensation, compared to lower-income families, in the reduction of the distance to travel for a marginal increase in the monthly stipend to pay and in the proportion of low-income students at the schools.

These systematic differences between socioeconomic groups may lead them to consider different types of schools, which also may make them skip different types of relevant alternatives. While parents from low-income groups are more conditioned to choose among the local supply, parents from higher GSE have more flexibility to consider farther schools with better academic quality. But whereas these results help to explain the type of schools that parents from different socioeconomic groups include in their choice set, it is insufficient to explain the individual and social consequences of skipping relevant alternatives. Further analysis must be done to explain the effects of making school decisions with an overwhelming amount of information.

6.2 Omission of more desirable schools

After estimating the preferences of parents for school characteristics, it is possible to use the utility function to test if parents are omitting more desirable schools. As a first approach, I will search for schools that have the same or better school attributes and are equal or closer than the one selected as the first choice in the application process. For this, I will consider the heterogeneous preferences based on the GSE of the block where the student lives. The utility function implies that the clone schools should have an equal or higher value of math test score and number of students per room as the school selected as most preferred in the application list, while also having equal or less

proportion of low-income students and a monthly stipend. For those who selected a secular school, clone schools can be secular or religious. But for those who selected religious schools, clone schools must impart religious teaching. Similarly, clone schools have one-gender or co-education for those parents who selected one-gender schools and only coeducational for those who did not select single-gendered schools.

This approach suggests that 20% of parents (16,180 of those who applied in 2019) are skipping schools that are more desirable and closer to their residence. Because of the results of the utility function, those parents that skip schools are omitting alternatives that would bring them more utility. This result suggests that there is a significant number of parents that cannot submit their preferred schools using an optimal strategy since they are omitting some more desirable and closer schools from the application list or relegating them to lower ranks. Focusing on the number of skipped schools, 9.8 out of those 20 percentage points are represented by parents that only omit one clone alternative. But another 6.3 pp. correspond to parents that skipped between two and four clone schools respectively. This indicates that 31.2% of those parents that skipped clone schools have more than one option that is better than the one selected as most preferred in the application list.

This evidence may be driven by the difficulty that parents face due to the overwhelming amount of information that they must consider while designing their application list. On average, parents have 37 schools closer than the one selected as first preference in the application list. If I consider those families whose first preference was at most 5 kilometers away, the number is reduced to 19. The high number of schools may lead parents to skip some alternatives to reduce the effort of comparing alternatives, evidencing the barrier that the costs of acquiring more information represent for making optimal decisions. For those parents whose preferred school is at most 5 kilometers away, the mean numbers of equal or closer school as the preferred one are 35 and 16 respectively. The above suggests that a possible reason why they are not choosing optimally is that they have twice the school supply that other parents have. Therefore, a bigger school supply may be detrimental if parents are not able to compare the whole offer, leading them to skip school to reduce the costs associated with gathering more information.

It is possible to argue that parents may have different capabilities of handling more information, depending on their socioeconomic level. However, the used approach indicates that income is not a factor that affects this capability. While 20.7% of parents from the lowest 50% of the socioeconomic distribution skip at least one clone school, 19.8% do so in the highest 30%. Focusing on those parents whose preferred school is at most 5 kilometers away, I also do not observe significant dissimilarities between the number of clone schools that parents from different GSE have. Those parents from the lowest 50% of the GSE distribution that skipped at least one clone school have on average 32 schools equal or closer that the most preferred school in the application list, while those who did not have on average 15 schools. In the case of parents from the highest 30% of the income distribution, the amount of school supply is 15 and 37 respectively. Since the inefficiency of the choices made by parents and the size of the school supply is not different between socioeconomic groups, I cannot ensure heterogeneity in the ability to manage information across socioeconomic groups.

While the previous way to measure how parents omit better schools is a reasonable approach that allows accounting for the preferences for school characteristics, figure 05 suggests that parents may consider more distant schools if they receive compensation in other attributes. Therefore, a second measurement is proposed. I consider those alternatives that may be farther from home but report parents a higher utility than the one selected as most preferred in the application list. The same heterogeneous utility functions are used to predict the utility level of each parent for each school. In this case, the proportion of parents that skip more desirable schools increase from 20% to 85.5%, where 34.4% of them would have to travel a longer distance (than the one selected as most preferred in the application list) to maximize their utility. Then again, it is not observed significant differences in the proportion of parents who skipped at least one better school among socioeconomic groups.

Comparing the variation of the school attributes that parents would observe if they considered their optimal choice instead of the one selected as most preferred in the admission list, some individual benefits of giving parents more information are observed. On average, families would prefer to apply to schools that are 700 meters closer to their location. While 33% of parents would have to increase the distance to travel, another 25% and 26% would be able to reduce the home-school distance in less and more than one kilometer respectively. For those parents that need to travel longer distances, they would receive an increase of 60 points in the school's math test score and a reduction of 20 pp. in the school's proportion of low-income students. On the other hand, parents who see their home-school distance reduced observe an increase of 24 points and a reduction of 6 pp. respectively. This is consistent with the idea that parents that must look farther need a higher compensation in other school attributes.

Figure 06: Change in the magnitude of the observed school attributes

School attribute	Socioeconomic distribution	C. 12.5	C. 25	C. 37.5	C. 50	Mean	C. 62.5	C. 75	C. 87.5
Distance	0%-50%	-1.958	-0.882	-0.314	0.000	-0.563	0.000	0.307	0.772
(Kilometers)	50%-70%	-2.352	-1.116	-0.467	-0.055	-0.734	0.000	0.274	0.736
	70%-100%	-3.028	-1.538	-0.655	-0.114	-0.993	0.000	0.188	0.826
Standardized Math Score	0%-50%	0.000	0.000	17.000	35.000	37.404	51.000	66.000	88.000
(Points)	50%-70%	-1.000	0.000	19.000	35.000	37.146	51.000	67.000	89.000
	70%-100%	-2.000	0.000	11.000	28.000	31.943	44.000	60.000	79.000
Monthly Stipend	0%-50%	0.000	0.000	0.000	0.000	8.058	0.000	0.000	52.430
(Thousands Of Pesos)	50%-70%	-18.840	0.000	0.000	0.000	16.516	22.499	43.500	70.600
	70%-100%	-11.403	0.000	0.000	9.367	25.451	35.284	63.056	81.900
% Priority Students	0%-50%	-32.086	-20.593	-12.847	-5.889	-8.483	0.000	0.248	11.433
(Percentage Points)	50%-70%	-33.614	-24.463	-16.715	-10.197	-11.584	-2.692	0.000	6.141
	70%-100%	-29.234	-20.266	-14.610	-9.627	-12.114	-3.564	0.000	0.000
Students Per Room	0%-50%	-2.000	0.000	1.000	3.000	3.816	5.000	7.000	11.000
(Students)	50%-70%	-2.000	0.000	1.000	2.000	3.483	5.000	7.000	10.000
	70%-100%	-4.000	-1.000	0.000	1.000	1.994	3.000	5.000	8.000
Religious School	0%-50%	0	0	0	0	0	1	1	1
(1 = Religious)	50%-70%	-1	0	0	0	0	0	1	1
	70%-100%	-1	0	0	0	0	0	0	1
Single Gender School	0%-50%	0	0	0	0	0	0	0	0
(1 = Single Gender)	50%-70%	0	0	0	0	-0	0	0	0
	70%-100%	0	0	0	0	-0	0	0	0

Vualues represent the difference between the atribute observed in the optimalschool and the one selected as most preferred in the admission process.

If the heterogeneous preferences lead to considering different type of schools, the types of skipped schools may also differ across socioeconomic groups. Figure 06 shows the difference, for each school attribute, between the value of the school that maximizes the utility and the one selected as most preferred in the application list. It is possible to observe that while 36.4% of the lowest 50% of the income distribution is willing to travel longer distances to access better schools, 29.3% of the highest 30% of the GSE distribution would also prefer to travel more. Those same groups would also experience an increase in 66.2 and 55.5 points in the observed math test score when considering the skipped school that maximizes their utility instead of their first choice in the application list. These results are as expected since parents from the highest socioeconomic group are already selecting schools that, on average, are 330 meters farther and have 24 extra points in the math test score.

Figure 06 also shows that, when parents do not skip relevant alternatives, higher socioeconomic groups experience a bigger reduction in the home-school distance to travel and the school's proportion of low-income students. But to access those alternatives, they must pay on average $25 thousand extra pesos. On the other hand, parents from the lower 50% of the income distribution have a smaller reduction in the distance to travel and an increase in the monthly stipend to pay. These marginal changes are compensated by a bigger increase in the observed math test score. The given results suggest that, while richer parents' biggest benefit of not skipping schools would be increasing the probability of experiencing a significant reduction in the home-school distance, the biggest benefit for poorer families would be preferring schools with higher academic quality.

The two previous approaches explore the hypothesis that parents are skipping better schools, in terms of utility maximization. I observe that the type of schools that parents skip are different among socioeconomic groups, where high-income families omit closer schools while poorer parents skip better academic quality schools. While I found that parents are omitting some relevant alternatives due to an enormous school supply from where to choose, including them in the application list does not mean that parents would be able to access those schools. Parents who apply to an over-demanded school could not be matched to a better alternative, since schools have capacity restrictions. Therefore, it is necessary to simulate the admission process with complete information to observe if there are some social benefits of parents being able to order their application list optimally.

7.0 Conclusions

Previous studies have pointed to high information costs and heterogeneous preferences for school attributes as factors that cast doubt on the efficiency of the School Choice system. Although one of the objectives of this system is to give households the freedom to choose the desired educational program, it has been suggested that preferences, household barriers, and the inhomogeneous distribution of school attributes could lead to socially undesired results. This research has sought to contribute to the School Choice literature by approximating the magnitude of the effects of choosing without complete information and under different preferences according to socioeconomic level.

Through the revealed preferences of applicants from pre-kindergarten to 1st grade in the city of Santiago, this study confirms the aversion of households to traveling long distances, paying high monthly stipends, and high proportions of low-income students in the schools. Preferences are also observed for schools with higher SIMCE scores and with religious education. The foregoing reaffirms

the idea that distance and access costs represent access barriers, which could reduce the real supply of schools to which certain households can have access.

By including interactions with the household's socioeconomic level, it is observed that there are heterogeneous preferences for certain school attributes. Distance act as a stronger barrier for families from the low part of the income distribution, conditioning them to take the nearby school supply into greater consideration. Given the great spatial correlation between academic quality and socioeconomic status of the neighborhood, it is possible to point out that such aversion would condition lower-income households to consider schools of higher academic quality than those with higher incomes. Together with the fact that higher-income households value academic quality more intensely, this would promote higher levels of socioeconomic segregation between schools.

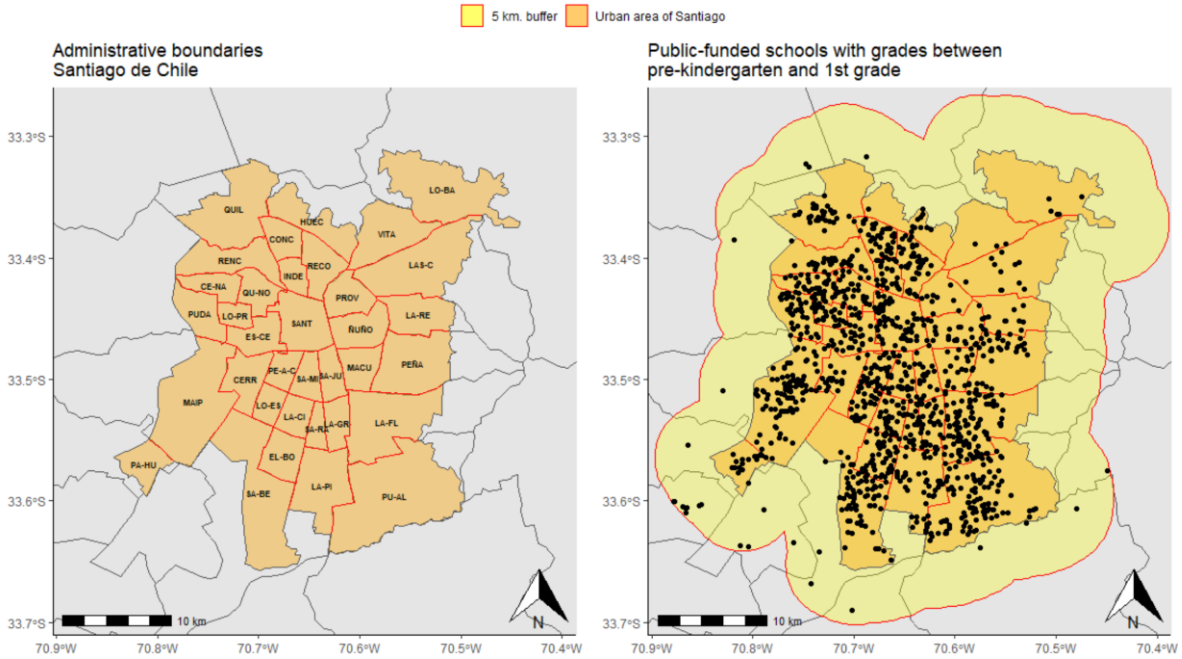
By constructing a utility function and estimating the predicted utility level of each household for each school, it is observed some parents are not selecting alternatives optimally. Considering clone schools that are equal or closer than the stated first preference, the approach suggests that 20% of parents are skipping schools that are more desirable and closer to their residence. In the case when the stated first preference is compared with the alternative that maximizes the utility level of each household, the proportion of parents that skip more desirable schools increase from 20% to 85.5%. By focusing on those parents whose stated first preference is at most 5 kilometers from the residential location, is it possible to suggest that the wide school supply in the area may be one of the reasons why parents skip better schools. Parents who did not maximize their utility level have, on average, twice the number of schools that those parents who did choose optimally.

Comparing the variation of the school attributes that parents would observe if they considered their optimal choice instead of the stated first preference, some individual benefits of giving parents more information are observed. On average, families would prefer to apply to schools that are 700 meters closer to their location. Since parents from all socioeconomic groups consider academic quality as a positive attribute, households consider alternatives with higher math test scores that the alternatives stated as most preferred in the application list. Parents would also experience a decrease in the observed proportion of low-income students enrolled at the schools. If parents do not have income restrictions or other barriers that limit their selection capabilities, parents will prefer alternatives with higher positive attributes and lower negative attributes to maximize their utility level. Therefore, identifying these barriers (income and information limits) is crucial to improve choice capabilities and provide equal opportunities in the school choice system.

Given the above, this research finds evidence that indicates that households are unable to optimally build their school's application list according to preferences. Finding schools that provide greater utility (which are usually associated with a higher math test score, lower proportion of low-income students, and a shorter distance to travel), it is observed that both heterogeneous preferences and lack of information influence the process of the school admission process. Therefore, it is important to study ways to reduce this information gap to improve the efficiency of the system.

8.0 Tables and plots

8.1 Administrative boundaries and school distribution



8.2 Description of specific variables of individuals, alternatives, and others

Interactive variables between alternatives and individuals		
Distance	Continuous	Euclidean distance (in kilometers) between a given household and school.
Individual-specific variables		
Priority student	Dummy	Indicates if the applicant student is a priority (1) or not (0). It is used as a proxy for the socioeconomic level of the household
Socioeconomic percentile (GSE)	Continuous	Socioeconomic percentile of the household block, average between blocks within a radius of 300 meters around the reported location.
Alternative-specific variables		
Math test score	Continuous	Score obtained by the school on the SIMCE standardized math test.
Monthly stipend	Continuous	Monthly cost of the copayment in subsidized private schools.
Proportion of low-income students	Continuous	Proportion of priority students who are enrolled in the school. It is used as a proxy for the socioeconomic level of the establishment.
Students per room	Continuous	Mean students per classroom.
Is Religious (1 = YES)	Dummy	Whether the school has a religious education (1) or not (0).
One Geneder (1 = YES)	Dummy	Indicate if the school admits only one gender (1) or both sexes (0).

9.0 References

Ajayi, K., & Sidibe, M. (2020). School choice under imperfect information. Economic Research Initiatives at Duke (ERID) Working Paper, (294).

Billingham, C. M., & Hunt, M. O. (2016). School racial composition and parental choice: New evidence on the preferences of white parents in the United States. Sociology of Education, 89(2), 99-117.

Bonal, X., Zancajo, A., & Scandurra, R. (2019). Residential segregation and school segregation of foreign students in Barcelona. Urban Studies, 56(15), 3251-3273.

Bonilla-Mejía, L., Bottan, N. L., & Ham, A. (2019). Information policies and higher education choices experimental evidence from Colombia. Journal of Behavioral and Experimental Economics, 83, 101468.

Bronfman, J. (2007). School Performance Evaluation under the Voucher System: The Case of Chile. The Wagner Review.

Burgess, S., Greaves, E., Vignoles, A., & Wilson, D. (2015). What parents want School preferences and school choice. The Economic Journal, 125(587), 1262-128.

Chumacero, R. A., Gómez, D., & Paredes, R. D. (2011). I would walk 500 miles (if it paid): Vouchers and school choice in Chile. Economics of Education Review, 30(5), 1103-1114.

Eyzaguirre, S., Hernando, A., Razmilic, S., & Blanco, N. (2019). ¿Qué explica las diferencias socioeconómicas en las preferencias escolares?. Centro de Estudios Públicos, N°514.

Dustan, A. (2013). Peer Networks and School Choice under Incomplete Information. University of California, Berkeley.

Elacqua, G., & Fabrega, R. (2004). El consumidor de la educación: El actor olvidado de la libre elección de escuelas en Chile. Santiago de Chile: PREAL.

Friedman, M. (1962). The Role of Government in Education. Milton Friedman, Capitalism and Freedom, Chicago: University of Chicago Press.

Gauri, V. (1999). School choice in Chile: Two decades of educational reform. University of Pittsburgh Pre.

Hastings, J., Kane, T. J., & Staiger, D. O. (2009). Heterogeneous preferences and the efficacy of public school choice. NBER Working Paper, 2145, 1-46.

Hofflinger, A., Gelber, D., & Cañas, S. T. (2020). School choice and parents' preferences for school attributes in Chile. Economics of Education Review, 74, 101946.

Ijaiya, Y. (1999). Effects of over-crowded classrooms on teacher-students interactions. Ilorin Journal of Education, 19, 1-11.

Khan, P., & Iqbal, M. (2012). Overcrowded classroom: A serious problem for teachers. University of Science and Information Technology, 49, 10162-10165.

Lopez, E., & Greenlee, A. (2016). An ex-ante analysis of housing location choices due to housing displacement: The case of Bristol Place. Applied Geography, 75, 156-175.

McEwan, P. J., & Carnoy, M. (2000). The effectiveness and efficiency of private schools in Chile's voucher system. Educational evaluation and policy analysis, 22(3), 213-239.

McEwan, P. J., Urquiola, M., Vegas, E., Fernandes, R., & Gallego, F. A. (2008). School choice, stratification, and information on school performance: Lessons from Chile [with comments]. Economia, 8(2), 1-42.

McFadden, D. (1977). Modelling the Choice of Residential Location, Discussion Paper, Cowles Foundation, Yale Universit

Neilson, C., Allende, C., & Gallego, F. (2019). Approximating the Equilibrium Effects of Informed School Choice (No. 628).

Neilson, C. (2013). Targeted vouchers, competition among schools, and the academic achievement of poor students. Working aper, Yale University.

Oosterbeek, H., Sóvágó, S., & van der Klaauw, B. (2021). Preference heterogeneity and school segregation. Journal of Public Economics, 197, 104400.

Owens, A. (2020). Unequal opportunity: School and neighborhood segregation in the USA. Race and Social Problems, 12(1), 29-41.

Santos, H., & Elacqua, G. (2016). Segregación socioeconómica escolar en Chile: elección de la escuela por los padres y un análisis contrafactual teórico.

Valant, J., & Weixler, L. B. (2020). Informing school-choosing families about their options: A field experiment from New Orleans. New Orleans, LA: Education Research Alliance for New Orleans.

Weiher, G. R., & Tedin, K. L. (2002). Does choice lead to racially distinctive schools? Charter schools and household preferences. Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management, 21(1), 79-92.