WORKING PAPER Nº IDB-WP-01565

Exodus to Public School: Parent Preferences for Public Schools in Peru

Gregory Elacqua
Nicolas Figueroa
Andrés Fontaine
Juan Margitic
Carolina Méndez

Inter-American Development Bank
Education Division

December 2023

Exodus to Public School: Parent Preferences for Public Schools in Peru

Gregory Elacqua
Nicolas Figueroa
Andrés Fontaine
Juan Margitic
Carolina Méndez

Inter-American Development Bank
Education Division

December 2023

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
Exodus to public school: parent preferences for public schools in Peru / Gregory Elacqua, Nicolas Figueroa, Andrés Fontaine, Juan Margitic, Carolina Méndez.
p. cm. — (IDB Working Paper Series ; 1565)
Includes bibliographical references.
1. Public schools-Peru. 2. School choice-Peru. 3. Schools closings-Social aspects-Peru. 4. Coronavirus infections-Social aspects-Peru. I. Elacqua, Gregory M., 1972- II. Figueroa, Nicolas. III. Fontaine, Andrés. IV. Margitic, Juan. V. Méndez, Carolina. VI. Inter-American Development. Education Division. VII. Series.
IDB-WP-1565

JEL Codes: A20, D12, D83, I28

Keywords: School Choice, Centralized Assignment Systems, Private-Public Schooling, COVID-19

http://www.iadb.org

Exodus to Public Schools:

Parent Preferences for Public Schools in Perú*

Gregory Elacqua† Nicolas Figueroa‡ Andrés Fontaine‡ Juan Margitic†

Carolina Méndez †

December 20, 2023

Abstract

Due to an unprecedented rise in demand, in 2020 the Peruvian Ministry of Education implemented a centralized assignment mechanism that allowed thousands of students at various levels of education to move from the private to the public sector. In this paper, we empirically explore the determinants of accepting a public school assignment and, subsequently, remaining in the public system. Specifically, we exploit the randomness in the assignment of students to new public schools to causally estimate the influence of distance on the decision to accept a public school placement, and we explore its role in the decision to remain there. We also provide insights into various determinants of parental preferences. Our findings reveal that families care about distance from home to the assigned public school as well as the relative academic and peer quality with respect to their school of origin. Parents weigh these factors differently based on their familiarity with them. Consequently, experiencing a new school environment can alter the significance of specific attributes when it comes time to decide whether to stay at the assigned school. These findings offer valuable insights into how governments can strengthen the supply of public schooling.

JEL Classification: **A20, D12, D83, I28**

Keywords: School Choice, Centralized Assignment Systems, Private-Public Schooling, COVID-19.

---

*Our gratitude goes to the Ministry of Education in Perú for providing the data used in this study. We thank Diana Vásquez Villanueva, Benjamín Madariaga, and Jadira Sánchez Córdova for their support in implementing the project and performing the analysis. We gratefully acknowledge the Inter-American Development Bank for funding this research. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent. The authors have no conflicts of interest or financial and material interests in the results. All errors are our own.

†Inter-American Development Bank

‡Pontificia Universidad Católica de Chile

1 Introduction

In 2020, Peru's education system was shocked by an unprecedented surge in demand for public school vacancies on the part of students looking to migrate from private schools. The exodus of over 100,000 students into the public sector was likely fueled by a combination of factors: extended school closures due to COVID-19, private schools' poor adaptation to online education, and the cost of paying for private education during an economic crisis. Notably, similar trends were also observed in other countries in the region (Elacqua et al., 2022).

In response, the Ministry of Education (MINEDU) designed a centralized assignment mechanism to enroll these students in public schools: the 2020 Exceptional Enrollment (*Matrícula Excepcional*; ME) process. This new process employed a Deferred Acceptance (DA) algorithm, though parents were unable to declare their preferences. Instead, pseudo-preferences were determined based solely on the distance of each school from the student's home. The closest school was considered the first preference, the second closest the second preference, and so forth. The DA algorithm assigned students to schools based on these preferences, randomizing assignments when schools were oversubscribed. Once the assignments were published, parents had the option of accepting or rejecting the assigned school.

We leverage this random assignment to gauge the impact of home-to-school distance on two outcomes: the probability of parents accepting the assignment and, once accepted, the likelihood of the student remaining in that school a year later. While the literature on parental choice in the context of education underscores distance as a pivotal factor in parental choices (Hastings et al., 2005; Jacob and Lefgren, 2007; Hofflinger et al., 2020; Harris and Larsen, 2017), our analysis also extends beyond this, shedding light on various determinants that influence the two outcomes of interest. Specifically, we assess different observable attributes, including school size, class size (measured by the student-to-teacher ratio), math test scores, peer demographics, and teacher experience. For families that moved to the public sector, our study evaluates how their perceptions of specific public school attributes differ after a year of direct experience. While we refrain from making causal inferences based on these attributes, our findings highlight factors that parents consider relevant and how these perceptions change with time and experience.

Our study reveals detailed patterns in parental choices, providing valuable insights that can inform and shape responsive policy measures aimed at enhancing the quality and attractiveness of public education. Specifically, our unique context allows to identify some of the dimensions of public education valued by parents who choose private education, which can help governments to dispel misconceptions about the public education system.[1] Strengthening public education has many potential benefits. Indeed, the latter functions as a safety net for all members of society (Singh et al., 2014); it is an adaptable, large-scale apparatus that all students can rely on in times of need (such as with COVID-19 or climate-change-induced schooling disruptions). It furthermore works an equalizer of disparate family backgrounds, which can reduce economic segregation (Murillo et al., 2018; Elacqua, 2012; Balarin, 2015).

The theoretical framework guiding our subsequent empirical analysis of parental public education preferences consists of a two-period model. In the first period, students are presented with the opportunity to either accept or reject their initial public school assignment. In the second period, one year after joining their assigned school, they determine whether or not to remain there. In this setting, parents observe an additional value of "experimenting" and going to a public school, since there is an option value. If they discover that the school is a good fit, they can continue attending it and, if not, they have the option of returning to the private sector if they can afford it. The model also explicitly takes into account that the decision to enroll in a public school and the decision to continue attending the latter may appear differently to an econometrician. When deciding whether to initially accept the offer, parents are focused on the long term, with school

[1] *Este País*, May 1^{st}, 2014

characteristics playing a crucial role given their potentially lasting impact. However, it is important to acknowledge the presence of selection bias in the decision of whether or not to stay on in the public school, as parents who accepted the initial offer may be more sensitive to school characteristics. The model effectively addresses key empirical findings, particularly the shifts in preferences related to essential attributes, such as math test scores and geographical distance.

Our findings reveal that the distance between school and home significantly influences the likelihood of parents accepting the assigned public school as well as affects their decision to continue at the same school. Specifically, for every additional kilometer increase in the distance to the school, the probability of accepting the assigned school decreases by 1.5 percentage points, while the probability of staying at that school decreases by 1.99 percentage points. This suggests that families prioritize convenience and accessibility in their decision-making process. The impact of this factor becomes even more pronounced in the decision to stay after an initial experience in the assigned school.

Furthermore, we find suggestive evidence that parents are more inclined to accept school assignments when the average level of parental education at the assigned school improves relative to the school of origin. This suggests that parents prefer schools with a more highly educated parent community, possibly considering it an indicator of the overall calibre of the establishment. Parents are also influenced by school quality, as measured by student achievement. Higher standardized test scores in math (relative to the school of origin) positively impact their likelihood of accepting the assigned public school. This suggests that parents use the previous school's attributes as a benchmark when making their decisions. In addition, we find that parents are more inclined to accept schools with more experienced teachers, a larger student body, and a lower student-teacher ratio. However, this last outcome does not hold up when more stringent fixed effects are included in the analysis.

It is not possible to establish causality with respect to the likelihood of students remaining in their assigned schools one year on. This is because the sample consists solely of parents who previously accepted the assignment. As mentioned, we observe that distance to school negatively impacts the probability of continued enrollment. Additionally, factors such as the demographics of peers (average level of parental education), school quality (math scores), and the number of students compared to their previous school also seem to play a significant role in this decision.

Additionally, we find that parents' experience of public schools can positively affect their perceptions of the public system, as they are able to more realistically compare their new public school to other establishments, both public and private. Once they move into the public system, some of the factors they initially considered crucial in their decision to accept the assigned school may become less significant in their subsequent decision to stay or leave. Thus, although these factors continue to play a role, their intensity or importance in influencing the decision may change over time. As with the first stage, relative peer demographics and school quality remain relevant. However, the importance of peer demographics (mean of parents' education) and distance to school increases relative to the first period of the analysis, while relative school quality (math scores) decreases in importance. Although we acknowledge that the sample of parents who decide to accept their school assignment is systematically different from parents in the first stage, this population of parents is nonetheless of interest to policymakers.

In the literature on family preferences for schools, several studies examine the factors that influence parents' choices. For instance, Hastings et al. (2005), Jacob and Lefgren (2007), and Hofflinger et al. (2020) explore parents' preferences for school characteristics such as academic quality, proximity, and religious instruction and find that parents from different socioeconomic backgrounds prioritize these factors differently. Specifically, those of lower socioeconomic status tend to prioritize proximity, while parents of higher socioeconomic status prioritize student achievement and religious education. Similar findings are reported by Abdulkadiroğlu et al. (2020), who observe that parents generally prefer schools with higher test scores.

With regard to revealed preferences, Harris and Larsen (2017) find that while parents in New Orleans prefer schools with higher test scores and other value-added measures, they also value non-academic factors such as after-school care, extracurricular activities, and proximity to home. Similarly, Glazerman and Dotter (2017) emphasize the significance of commuting distance, school demographics, and academic indicators in school choice, highlighting the heterogeneity of preferences across parent demographics.

Our study aligns with previous research findings. For example, we also find that families who were previously in the private sector tend to prefer higher-performing public schools with favorable peer demographics and proximity to their residence, consistent with Jacob and Lefgren (2007) and Burgess et al. (2015). Our results also support the observations of Harris and Larsen (2017) regarding parents' consideration of both academic and non-academic attributes in school selection, as well as the findings of Glazerman and Dotter (2017) concerning the role of commuting distance, peer demographics, and academic indicators.

Broadly, we contribute to two different strands of the literature. First, we add to the body of work that exploits information on parental preferences and the randomization induced by centralized assignment mechanisms to better understand school choice (Abdulkadiroğlu et al. (2017), Glazerman and Dotter (2017), Kapor et al. (2016), Agarwal and Somaini (2018)). However, our study focuses on a unique case where preferences for the randomization algorithm implemented were centrally imputed based on distance, contrasting with scenarios in which parents have the ability to submit their own choices.

Second, we build on the literature showing that parental preferences for public schools encompass multiple dimensions, including academic quality, proximity, socioeconomic composition, and non-academic factors (Elacqua et al. 2006). Our examination of preferences through a natural experiment makes it possible to capture the perspectives of a large group parents who transitioned from private to public schools. This unique perspective may prove valuable for governments striving to improve public education. Indeed, these families, primarily from middle- and low-income backgrounds, represent a broader population that has shifted towards the public sector in recent decades. Better understanding how these parents view the public education system and the educational aspects they value is essential for governments seeking to formulate and implement effective policies to strengthen public schooling.

Our research provides insights into parents' preferences based on observed behavior rather than self-reported data. In general, the stated preference literature (Schneider et al., 1998; Kleitz et al., 2000) suggests that parents often prioritize teacher quality, high test scores, educational quality, class size, and safety when making school choices. However, the specific nature and extent of these preferences may vary depending on the survey question and the demographic characteristics of the parents.

This paper is structured as follows. Section 2 describes the context and events that led to the 2020 ME process, as well as the data used in our study. Section 3 introduces the theoretical model that guides our empirical analysis. Section 4 presents our empirical analysis, including the identification strategy employed, the main results, and a non-linear analysis. Finally, Section 5 discusses our findings and concludes.

2 Institutional Background

The COVID-19 pandemic led governments throughout Latin America to close schools due to the increased risk of contagion and a lack of available vaccines. Countries in the region shut down schools for an average of 158 days, in stark contrast with the OECD average of 57 days.[2] There were marked differences in schools' and students' capacity to adapt to this new situation. Schools faced critical constraints such as the availability of technological resources as well as the ability of the teachers to shift to remote instruction.[3] Students meanwhile were limited by the resources available at home to participate in online instruc-

[2]Source: Authors' calculations based on UNICEF data for the period from 11/3/2020 to 2/2/2021.
[3]Infobae, July 28 2022

tion, such as technological devices, internet connection, a physical space to study, etc. (Abizanda et al., 2022).

The pandemic had severe consequences in Peru. The country had the world's highest COVID-19 death rate, far greater than any of its neighbors and twice that of the United States. According to official records, COVID-19 caused nearly 6,470 deaths for every million people.[4] In addition, unemployment increased considerably. According to the national statistics agency (*Instituto Nacional de Estadística e Informática* [INEI]), during the second quarter of 2020, employment among the worst-affected groups, informal employees and independent workers, decreased by 60% and 35%, respectively. In absolute terms, this meant an annual loss of approximately 4.8 million jobs in the informal segment of the labor market.

In the context of education, private schools were reticent to lower their fees[5] and even threatened to expel students in extreme cases.[6] The large numbers of students seeking to transfer from private to public schools was likely a result of this reluctance on the part of private schools, coupled with economic hardship brought about by the pandemic. In response, the government designed a system that could accommodate a large number of transfer requests while preserving the principles of equity and efficiency. Specficially, the MINEDU, in collaboration with the Education Division at the Inter-American Development Bank, implemented a centralized assignment mechanism using a Deferred Acceptance (DA) algorithm to match each student with the closest available vacancy to their residence.

2.1 Exceptional Enrollment Program in Peru - ME 2020

The MINEDU launched the Exceptional Enrollment Program (in Spanish, the "*Matrícula Excepcional*," or ME) on May 22, 2020. The program was designed to ensure educational access during these extraordinary circumstances, when a surge occurred in demand for public education. Families could submit their applications through a web portal, where

[4] Our World in Data

[5] *El Comercio* newspaper, April 26, 2020

[6] *Revista Matg*, September 25, 2020, *Otras Voces en Educación*, May 28, 2020

they were asked to include details such as the student's grade, age, and address, which could either be their residential address or another address.

The ME utilized a Deferred Acceptance algorithm, which did not, however, allow parents to specify their preferred schools. Instead, preferences were automatically determined based on the distance from the student's home to potential schools within a 5-kilometer radius. The closest school was designated as the first preference, followed by the next closest as the second preference, and so on. A sample of these auto-generated preferences is illustrated in Table 1. The algorithm gave priority to students with siblings attending a specific public school and students with special educational needs arising from disabilities.

After registration, families could view their school assignments from June 7th to June 21st. They were free to either accept or decline their designated public school. If a family declined because they had registered an incorrect address, they could update the information and reapply. The algorithmic procedure ran three times to accommodate students, particularly those who initially turned down their placements due to address inaccuracies.

The program categorized requests into two types: "first entry" for newcomers to the Peruvian education system and "transfer" for those moving from a private school to a public school or between public schools. Additionally, the MINEDU classified vacancies into regular, extended, and virtual. Regular and extended slots catered to traditional, in-person learning, while virtual slots supported a blended learning model, combining self-paced study at home with remote guidance from teachers. It is important to note that all instruction was conducted virtually in 2020 and 2021.

2.2 Data

Our starting sample includes all 125,295 students who participated in the ME 2020 process, mainly consisting of pupils enrolled in private schools in urban settings. Parents' decision to submit an application was likely a combined response to extended school closures, private schools' lack of adaptability in transitioning to online education, and the continued

Table 1: Example of Pseudo-Preferences

Student	Ranking	School	Distance
A	1	ABC	0.1 km
A	2	DEF	1.2 km
A	3	GHI	3.0 km
A	4	JKL	4.8 km
B	1	DEF	1.3 km
B	2	ABC	3.4 km
B	3	JKL	5.0 km

requirement of paying private school tuition fees.

We use data provided by the MINEDU on students' background, their parents' decisions, and schools' characteristics to investigate the possible factors that determine family decisions. We obtained individual records of students' enrollment status and teacher characteristics from 2019 to 2021 from the Information System to Support the Management of the Educational Institution (*Sistema de Información de Apoyo a la Gestión de la Institución Educativa*; [SIAGIE]) database. The information on average school characteristics comes from multiple databases: measures of parental education and schools' standardized academic performance from 2014 to 2019 from the Census Evaluation of Students (*Evaluacion Censal de Estudiantes* [ECE], a nation-wide assessment) and data on staff and facilities from the School Directory (*Padron Escolar*) as well as the School Census (*Censo Escolar*) from 2014 to 2021.

2.3 Sample Definition

We are interested in analyzing two distinct stages of the school allocation process. The first stage involves parents either accepting or rejecting the assigned public school for their children. The second stage consists of families deciding whether to remain or leave the previously assigned and accepted public school. During both stages, we exclude students who participated in the second and third rounds of school assignments, given that they faced different pools of potential schools for assignment and some of their preferences were inaccurately presumed due to incorrect address information. Furthermore, we focus our analysis exclusively on students who were previously attending private schools, as our aim is to examine a population that had initially opted for the private over the public sector.[7]

To ensure an authentic random distribution in our process, we excluded clusters where each student was assigned to their preferred school. These situations indicate that randomization had not been necessary as everyone received their first choice. Conversely, students being assigned to their second choice or lower indicates that randomization was indeed employed. We identified these clusters using two methods. First, we grouped students into a single cluster if they were from the same district, had the same educational level, and were in the same grade[8]. For a more rigorous identification, we formed clusters of students with the same top two school preferences on their lists.

In the second stage, we exclude students who did not accept their assignments from the first round[9] as they did not gain firsthand experience of the public school environment. Indeed, the goal of our second stage analysis is to identify the factors that determine a student's decision to stay in a public school after having had firsthand experience there.

In addition, we further restrict our sample to students in the second stage to students who attended a school in 2021 within a 50-kilometer radius of the school to which they were assigned in the 2020 ME process. Indeed, students may switch schools for reasons unrelated to the schools' characteristics, such as family relocation. Our analysis is limited to students who, in 2021, were enrolled in a school located close to their initial assignment. Additionally, we note that 3,588 students completed their education between 2020 and 2021 and are therefore not included in our sample.[10]

We define *accepting* a school assignment as a student accepting and virtually attending their assigned school. This is because some students accepted their designated school yet

[7] $8,487$ students, or 6.8% of our sample, were enrolled in public schools.

[8] We use the term "educational level" to refer to nursery, primary, or secondary levels. Within each level, the term "grades" distinguishes a student's academic progress, such as grades 1, 2, 3, etc.

[9] The proportion of students who accepted their first-round assignment is 82.9%.

[10] Additionally, due to incomplete administrative records, we were unable to locate 2,017 students in 2021 or 2020. Consequently, these students have been excluded from the second stage of our analysis given that we can't explain their absence and modelling their participation would require further assumptions.

did not proceed with enrollment.[11] In these cases, we view these students as having *de facto* rejected their assignment, and re-code their responses in the first stage. and explore its role in the decision to remain as being registered at a different school after the ME allocation occurred. For example, if a student accepts a placement from the ME2020 process but is later found registered at another school in September, we classify that student as "treated." We infer that this student spent a few months at the assigned school, after potentially starting there in June, and then decided to switch in September.[12]

After applying our sampling criteria, we retained 63.23% (or 79,230 students, with 38,401 of them grouped by their top two school preferences) of the original pool of students in the first stage of the 2020 ME process.[13] In the second stage, our sample comprises 51.81% (or 64,916 students, with 31,637 grouped by their top two school preferences) of the original population.[14] Figure 1 illustrates the progression of students (who come from the private sector) from the first to the second stage. A horizontal dotted line divides the sample into two stages. Green arrows represent students' decisions to pursue education in the schools allocated during the ME2020 process. In contrast, red arrows symbolize their choices to decline or leave these placements. The first row of arrows indicates the initial decisions to either accept or decline the ME2020 school assignments. The second row shows the students who accepted their placement but never actually enrolled in the assigned schools. We label these *"de-facto"* rejections. The third and fourth rows map out student decisions to transfer out of their assigned schools, either after spending a few

[11] We define these students as those who accepted their assignment but never formally enrolled in these schools. In Figure 1, they number 6,323 (6,245 + 78) out of the 71,916 students who initially accepted their assignment.

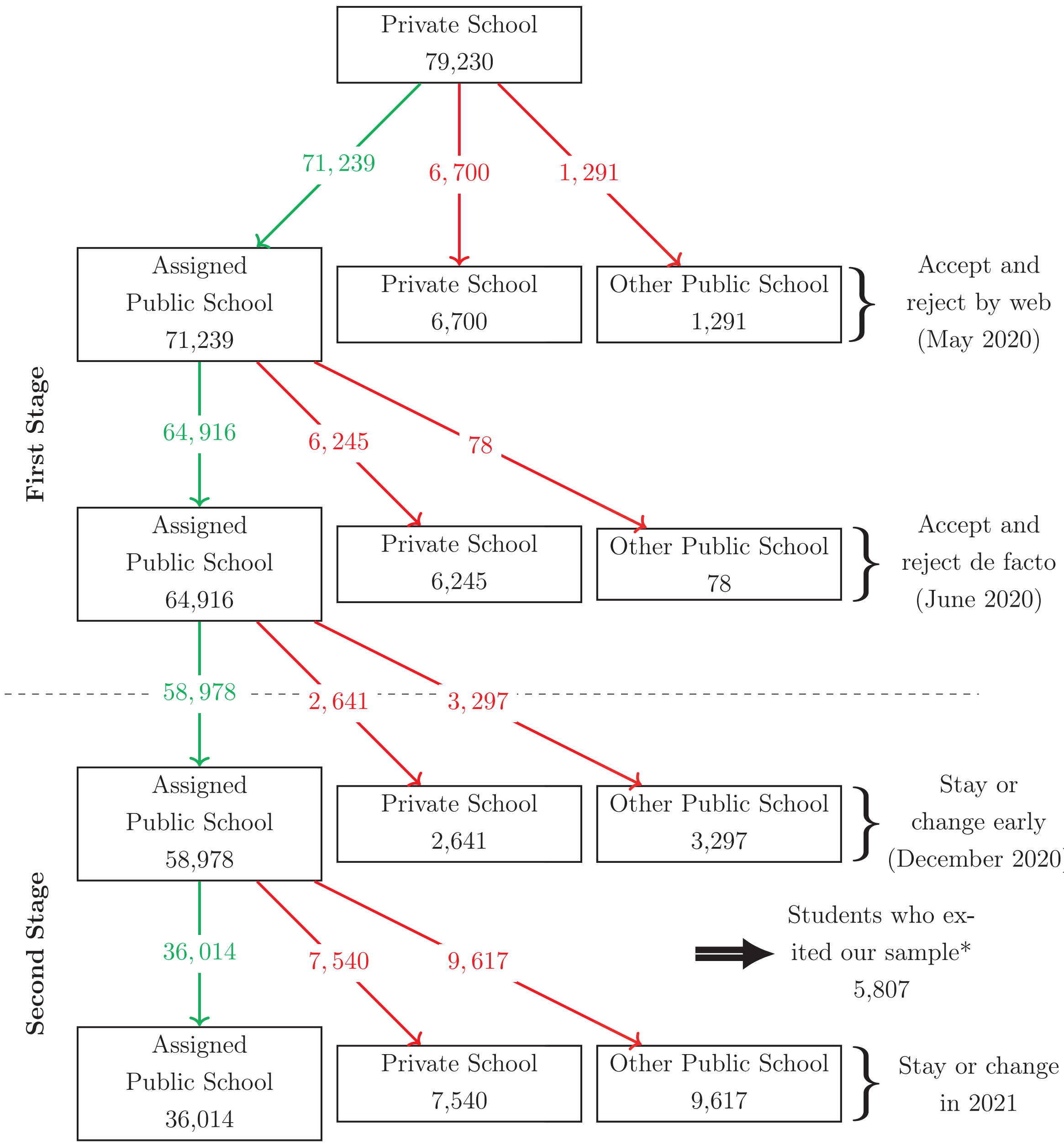
[12] In Figure 1, this group accounts for the 2,641 students who migrated to private schools in 2020 and the 3,297 students who migrated to a different public school, out of a total population of 64,916 students who were effectively "treated."

[13] Please refer to Table 9 in Appendix A.2.1 for a breakdown by educational level and grade.

[14] It is important to note a distinction between our regression samples and the figures we report. This disparity arises for two main reasons. First, our preferred analysis approach involves breaking down the samples based on district, grade, and education level. In doing so, we found that many subsets lacked sufficient variation in decisions to accept or reject school assignments, resulting in approximately 5,000 students being excluded from our study. Second, we faced challenges due to missing data on certain analytical factors for some students. Though, this only impacted a relatively small portion of our overall sample (approximately 3,000 students).

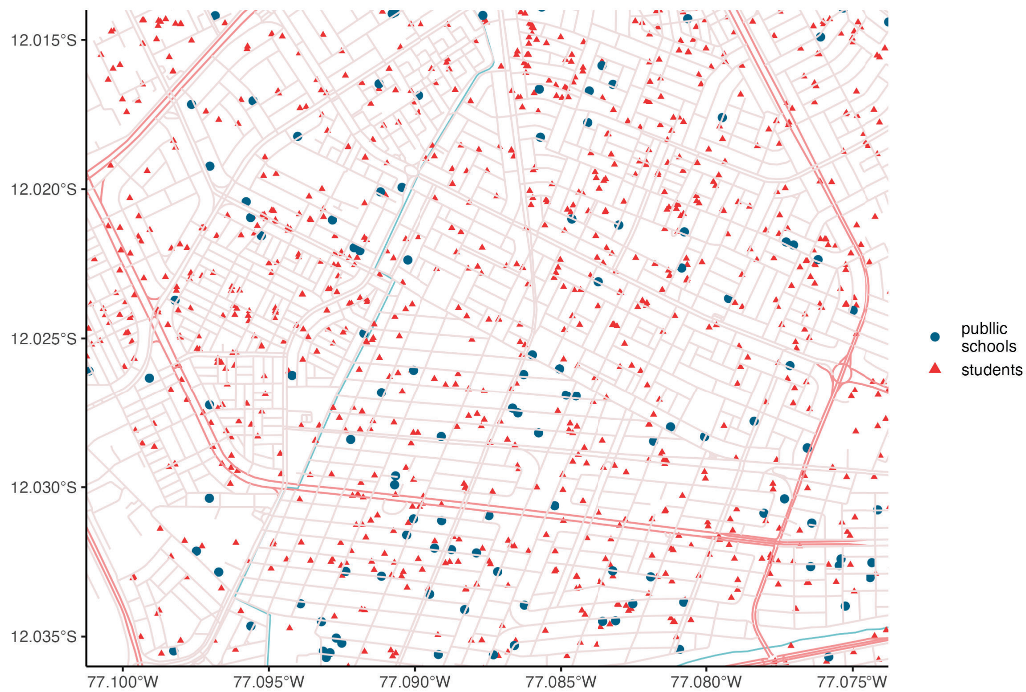
months there in 2020 (third row) or in subsequent decisions made by March 2021 (fourth row). An example of the geographical distribution of students and schools included in our study is shown in Figure 2.

Figure 1: Student flow before and after 2020 ME



*Among these students, 1,143 were excluded from our analysis as they had a distance greater than 50 km between their assigned school and the school they enrolled in for 2021. Additionally, 2,647 students graduated in 2020, and records for 2,017 students are incomplete or missing.

Figure 2: Geographic Distribution of Students and Schools Participating in the ME2020



(a) Note: The higher color concentration indicates a higher density of students and schools.The area shown is a neighborhood in metropolitan Lima

2.4 Sample Description

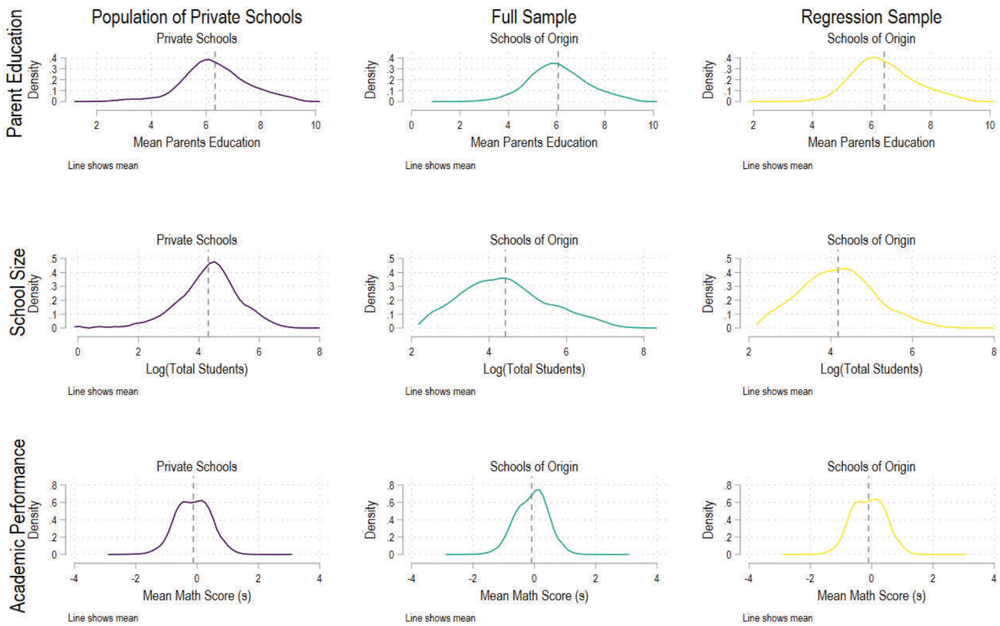
In this section, we discuss in detail the characteristics of the schools that participated in the ME 2020 process.

We begin by comparing the private schools of origin of the students who participated in the ME 2020 process (which we term the "Full Sample") to all the private schools in the Peruvian school registry (referred to as the "population of private schools in Peru"). The former sample is further compared to the schools included in our final dataset, termed the "Regression Sample." We compare these three samples by plotting the distributions of school characteristics in Figure 3 and Panel A of Figure 4. They demonstrate that, even after applying our selection criteria, our sample remains representative of the broader school population.

We then repeat this analysis focusing on public schools. We contrast the attributes of all public schools in the registry to the schools assigned to students during the ME 2020 process ("Full Sample") and those present in our "Regression Sample." As illustrated in Figure 5 and Panel B of Figure 4, subtle variations emerge in the distributions. The assigned schools exhibit higher levels of parental education, larger school populations, and a lower prevalence of establishments with a very low socioeconomic index compared to public schools in the broader population. This observation underscores a trend of students not being placed in the most disadvantaged schools, which may indicate a degree of parental selectivity.

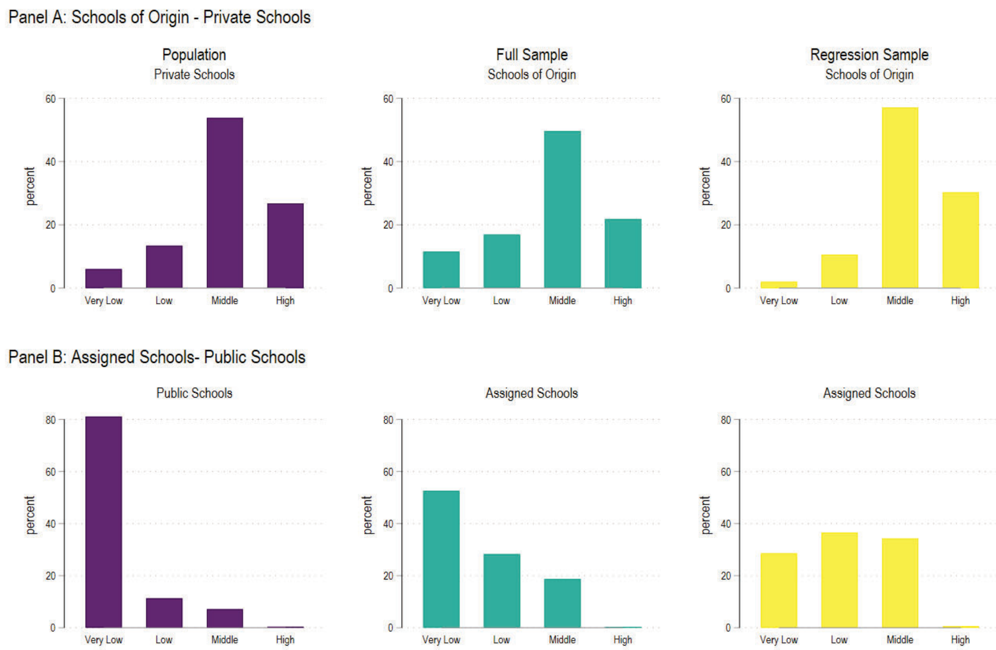
Next, narrowing our analysis to the regression sample, we examine the evolution of covariates between stages (Table 2). Contrary to initial expectations, there are no substantial changes observed between the stages. This suggests a notable stability in the factors measured across both stages of the analysis. Along the same lines, Table 3 shows the evolution of the differences in the covariates between the assigned school and the school of origin. Here, we observe more significant changes. For instance, the disparity in school math scores is higher among students who accepted their assignments, indicating that

Figure 3: Comparative Distribution of Characteristics in Private Schools among the Population, Full Sample, and Regression Sample



those who accepted their public school assignments went to schools with higher average test scores in math compared to their schools of origin. Indeed, this disparity is even more pronounced among students who decided to stay in the public system. A similar pattern emerges with regard to parental education: while the assigned schools have lower average levels of parental educational compared to the schools of origin, this gap narrows when

Figure 4: Distributions of Socioeconomic Index in Private and Public Schools.



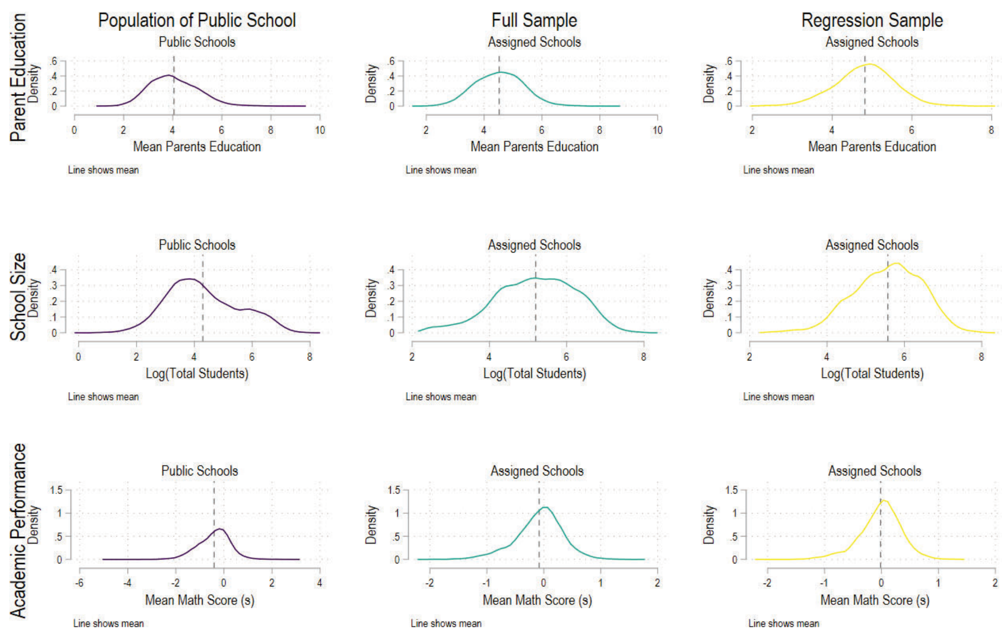
the sample is limited to public schools where students both accepted and stayed.[15] Additionally, the assigned schools where students chose to accept and stay tend to have larger student populations than the schools of origin.

Interestingly, our analysis reveals that while there are minimal alterations in the covariates between the different stages, pronounced disparities do emerge in terms of the differences between assigned schools and schools of origin. This suggests that parents

[15] Parental education is measured as a categorical variable and represents different levels of educational attainment by parents. These levels are averaged at the school level based on data from the ECE dataset. Below is an explanation of what each numerical value represents:

1. No education
2. Incomplete primary education
3. Complete primary education
4. Incomplete secondary education
5. Complete secondary education
6. Incomplete non-university higher education
7. Complete non-university higher education
8. Incomplete university education
9. Complete university education
10. Continuing education post-university

Figure 5: Comparative Distribution of Characteristics in Public Schools among the Population, Full Sample, and Regression Sample



make decisions based on their prior experiences with a school, and this experience may significantly change following the initial school assignment.

Table 2: Covariates by Stages

	Sample			Accept			Stay		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Distance (km)	76343	2.15	1.40	50069	2.10	1.40	35052.00	2.08	1.40
STR	76324	22.48	4.94	50056	22.57	4.97	35044.00	22.75	4.91
Mean Math sc (s)	60515	0.00	0.31	42596	-0.00	0.32	30551.00	0.01	0.31
Mean Parent Edu.	60508	4.93	0.62	42591	4.93	0.62	30551.00	4.95	0.60
Log(Num. of Students)	76324	6.03	0.75	50056	6.10	0.73	35044.00	6.16	0.70
Prop. Novice Teachers	75913	0.14	0.17	49797	0.14	0.17	34852.00	0.13	0.16
Cost School of Origin	75495	236.37	195.39	49488	229.18	186.16	34605.00	220.32	179.18

Table 3: Difference in Covariates by Stages

	Sample			Accept			Stay		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Dif. Distance (km)	75047	-0.07	4.06	49019	-0.16	4.16	34441.00	-0.15	4.15
Dif. STR	74795	5.29	9.28	49055	5.54	9.33	34338.00	5.82	9.35
Dif. Mean Math sc (s)	56471	0.08	0.56	39183	0.12	0.55	28073.00	0.15	0.54
Dif. Mean Parent Edu.	55422	-1.45	1.01	38338	-1.38	0.99	27419.00	-1.30	0.95
Log(Ratio Num St.)	75096	1.22	1.05	49231	1.30	1.05	34462.00	1.38	1.04
Dif. Prop Novice Teachers	72362	-0.20	0.38	47427	-0.20	0.38	33174.00	-0.20	0.38

(a) Note: Differences are calculated by subtracting the value of each variable of the school of origin from that of the assigned school.

3 Theoretical Framework

We introduce a two-period model that allows us to rationalize the primary empirical findings in Section 4.2. In the first period, students/families decide whether to switch to the public sector based on the school they are offered. In this initial decision, students' preferences for the offered school are imperfectly informed. While they have some information about certain characteristics (e.g., distance), they lack knowledge about the quality of the match with the school. After attending the school, they learn about this match component and decide whether to stay or leave. Students fully internalize the "experimentation value" of switching to a public school: they can remain if they like it or return to a private school if they do not.

The model effectively explains certain empirical observations, especially the shift in the emphasis placed on key attributes such as math test scores and distance. This can be attributed to two competing effects: initially, when a student is deciding whether to enroll in a public school, the school's characteristics play a pivotal role since they will influence the student over two periods. This decision also carries an option value, as students will continue their education at the school only if they find it sufficiently appealing. Conversely, when students are deciding whether to continue at the school, the decision is made by a specific subset of the initial population: those who initially chose to try the public school. This group may place a higher emphasis on these school characteristics. The magnitude of these competing effects will determine whether school attributes exercise a greater influence over the decisions in the first or second stage.

In the first period, a student has two options: a public school (PU) and a private school (PR), taking into account the present period and expectations for the second period. The student's payoff for attending the private school is $V_{PR}^{(1)} + \varepsilon$ in the first period and $V_{PR}^{(2)} + \eta$ in the second. Here, we assume that $V_{PR}^{(2)} > V_{PR}^{(1)}$ to reflect the fact that households are experiencing a crisis in the first period due to COVID-19, which reduces their valuation for the private school, either because of economic difficulties or a decrease in service quality.

Additionally, ε and η are iid idiosyncratic shocks distributed according to G with a mean of 0. These shocks represent variations in the level of satisfaction with the current school (ε) and the strength of recovery after the crisis (η). Agents know the realization of ε before making their decision at $t = 1$ and the realization of η before their decision at $t = 2$. They do not know in advance what their economic situation will be after the pandemic or how well their original school will recover from the shock.

Students face significant uncertainty about the public school (PU) option since it is unknown to them. We assume that the valuation for the public option is αV_{PU}, where V_{PU} is an objective measure of school attractiveness that captures observable attributes such as distance, parental education, and schools' results on standardized tests. Its empirical counterpart, V_{PU}, is represented by the vector of observable attributes of a given school.

Meanwhile, α is an idiosyncratic shock that represents the quality of the school-student match, something not observable by the econometrician and also unknown to students before attending a school. Students are initially uninformed about α, which encompasses, among other things, the relationship with classmates, teachers, and general satisfaction with the learning environment. We assume two types of students, denoted as $i = \{1, 2\}$ in proportions μ_0 and $1 - \mu_0$. The distribution of α is specified as F_i on [1,K], with $K < \infty$. We assume that $F_1 \succeq_{FOSD} F_2$, signifying that a fraction μ_0 of the population is more likely to have a positive experience in the public school system. Students are aware of the group to which they belong, indicating that some of them are more willing to try out the public sector. We assume rational expectations, meaning that students are aware of the likelihood that they will have a good match with the public school, and there is an experimentation motive since attending a public school is necessary to dispel misconceptions.

The following assumption captures the idea that a bad match with public schools ($\alpha = 1$) makes it undesirable even in the first period, but a good match ($\alpha = K$) makes it the best option even in the second period, after the crisis has passed.

Assumption 3.0.1. *We assume that* $V_{PU} < V_{PR}^{(1)} < V_{PR}^{(2)} < KV_{PU}$.

In the first period, students choose whether to accept or reject the public school offered by the government. They do so knowing their valuations for the private school ($V_{PR}^{(1)} + \varepsilon$ in the first period and the expectation of $V_{PR}^{(2)}$) in the second period), but with only an uncertain expectation of their valuation for the public school option ($\mathbb{E}_{F_i}(\alpha)V_{PU}$). Additionally, there is an "exploration value" associated with the public school option because the parameter α becomes known after the first period if the student experiments with the public school. Therefore, in $t = 1$, a student of type i accepts the public school offered if:

$$\underbrace{\int_1^K \alpha V_{PU} dF_i(\alpha)}_{\text{First period expected payoff}} + \underbrace{\int_1^K max\left\{\alpha V_{PU}, V_{PR}^{(2)}\right\} dF_i(\alpha)}_{\text{Second period expected payoff}} \geq \underbrace{V_{PR}^{(1)} + V_{PR}^{(2)} + \varepsilon}_{\text{Private school payoff}} \quad (1)$$

where ε is distributed according to G and known by the student. In other words, a student chooses to experiment if and only if the expected payoff in the first period plus the expected payoff in the second period, once α is known, surpasses the secure option of remaining in the private sector. Experimentation holds intrinsic value due to its option value, as the student will continue in the public sector if $\alpha \geq \frac{V_{PR}^{(2)}}{V_{PU}}$, and return to the private sector otherwise.

Lemma 3.0.1. *In $t = 1$, a student of type i accepts the public school if and only if:*

$$H_i(V_{PU}) \equiv V_{PU} \cdot \left[\mathrm{E}_i(\alpha) + \int_{\frac{V_{PR}^{(2)}}{V_{PU}}}^K \alpha dF_i(\alpha)\right] - \left[V_{PR}^{(1)} + V_{PR}^{(2)} \cdot (1 - F_i\left(\frac{V_{PR}^{(2)}}{V_{PU}}\right))\right] \geq \varepsilon$$

In essence, for each type of student, there are two threshold values for ε, beyond which the student chooses not to accept the public school assignment. Therefore, the proportion of students that go to public schools is given by:

$$\mathbb{P}(try) = \mu_0 G(H_1(V_{PU})) + (1 - \mu_0)G(H_2(V_{PU}))$$

where $G(H_1(V_{PU})) > G(H_2(V_{PU}))$

Proof. Direct from algebraic manipulation and the fact that $F_1 \succeq_{FOSD} F_2$. $\square$

In $t = 2$, students who tried out a public school choose whether to stay in the public system or not. In this period, they already know their match with the public school (α)

and their valuation of going back to the private sector ($V_{PR}^{(2)} + \eta$). Thus, a student stays in their public school if:

$$\alpha V_{PU} \geq V_{PR}^{(2)} + \eta \tag{2}$$

Analogously to the previous lemma, characterize the proportion of students that stay in public schools.

Lemma 3.0.2. *At $t = 2$, a student stays in the public sector if:*

$$L(\alpha V_{PU}) \equiv \alpha V_{PU} - V_{PR}^{(2)} \geq \eta$$

In simpler terms, there exists a specific threshold below which a student chooses to continue in the public school. Therefore, the proportion of students who remain in the public system is determined by:

$$\mathbb{P}(stay) = \mu_1 \int_{\alpha=1}^{K} G(L(\alpha V_{PU}))dF_1(\alpha) + (1-\mu_1) \int_{\alpha=1}^{K} G(L(\alpha V_{PU}))dF_2(\alpha)$$

where $\mu_1 = \mathbb{P}(type1|try) = \frac{\mu_0 G(H_1(V_{PU}))}{\mu_0 G(H_1(V_{PU}))+(1-\mu_0)G(H_2(V_{PU}))} > \mu_0$

Proof. Direct. □

By itself, the previous lemma has two important empirical implications. First, as V_{PU} increases, students are more likely to go to the public school to which they were assigned. Second, students who accept this option are more likely to belong to group 1, which has a higher expected match with the public school system. In other words, students of type 1, who are more inclined to prefer the public school option, are over-represented in the sample at $t = 2$, that is, $\mu_1 > \mu_0$. This phenomenon occurs because at $t = 1$, before they ascertain their actual match with the public system, they are more optimistic and thus more willing to experiment during the crisis. This *selection bias*, characterized by the difference $\mu_1 - \mu_0$, plays a pivotal role in our results.

We are interested in analyzing the sensitivity of parents' decisions to school characteristics, both at $t = 1$ and $t = 2$, and the changes in this sensitivity. In the model, this is reflected in the quantities $\frac{\partial \mathbb{P}(try)}{\partial V_{PU}}$ and $\frac{\partial \mathbb{P}(stay)}{\partial V_{PU}}$, which indicate how a student's (expected) decision changes when offered a marginally better school. The empirical counterpart of these quantities is given by the parameter β_i in the regressions, which indicates the increased likelihood of an option given a change in the observable characteristics of a school.

As expected, the likelihood of trying the public school at $t = 1$ and staying in at $t = 2$ are increasing in V_{PU}. Interestingly, the sensibility to changes in V_{PU} can be greater or smaller in $t = 2$, depending on the strength of the selection bias. This selection bias is composed of two terms. On the one hand, for any V_{PU}, we have that $\mu_1 > \mu_0$. That is, parents already enrolled in the public school system are more sensitive to school quality. On the other, as V_{PU} increases, the magnitude of this bias $(\mu_1 - \mu_0)$ also changes.

Lemma 3.0.3. *Suppose that G is a uniform distribution on an interval $[-a, a]$. The sensibility in the second period will be greater if:*

$$(\mu_1 - \mu_0)(\bar{\alpha}_1 - \bar{\alpha}_2) + \lambda(V_{PU}) > \left[(1 - \mu_0)\int_{\alpha=\frac{V_{PR}^{(2)}}{V_{PU}}}^{K} \alpha \cdot dF_2(\alpha) + \mu_0 \int_{\alpha=\frac{V_{PR}^{(2)}}{V_{PU}}}^{K} \alpha \cdot dF_1(\alpha)\right] \quad (3)$$

where $\lambda(V_{PU}) = \frac{\partial \mu_1}{\partial V_{PU}} \left[\int_{\alpha=1}^{K} G(L(\alpha V_{PU}))dF_1(\alpha) - \int_{\alpha=1}^{K} G(L(\alpha V_{PU}))dF_2(\alpha)\right]$

The proof can be found in the appendix A.1

In condition (3), the left-hand side quantifies the magnitude of the selection effect. As "optimistic" students, with higher values of α, are over-represented at $t = 2$, the importance of the actual observable school characteristics, V_{PU}, becomes more acute. On the other hand, the right-hand side of (3) represents the *horizon effect*. At $t = 1$, V_{PU} assumes greater importance because it has the potential to influence the student over both periods (always at $t = 1$, and with some probability at $t = 2$).

When the selection effect is more pronounced, we can expect a higher level of sensitivity in the second stage. Conversely, when the horizon effect is purely mechanical in nature, as in the first stage, students consider two periods instead of just one. If the selection effect prevails, we should observe a greater sensitivity in the second stage.

It is important to note that $\lambda(V_{PU}) > 0$ if $\frac{H_1'(V_{PU})}{H_2'(V_{PU})} > \frac{G(H_1(V_{PU})}{G(H_2(V_{PU}))}$. In such a scenario, the selection effect intensifies as V_{PU} increases. This is true as long as long as H_1 is more elastic than H_2,[16], which is to say V_{PU} induces a bigger percentage change in the school

[16]That is, $x\frac{H_1'(x)}{H_1(x)} > x\frac{H_1'(x)}{H_1(x)}$.

valuation for group 1 than for group 2.

Additionally, $\mu_1 > \mu_0$ since, in the second stage, only the students who accepted their assignment in the first stage are present. Given that F_1 exhibits first-order stochastic dominance over F_2, students who accepted are more likely to belong to type 1. We refer to this as the selection effect.

The relative strength of the selection and horizon effects determines which of the two decisions – to try the public school or to stay enrolled in it – is more sensitive to the actual characteristics of the school offered. This insight also explains why certain school attributes (e.g., distance and peer demographics) become even more pivotal for students' decisions *after* they have already been enrolled in the public school for a year. The findings in section 4.1.3 align with a strong selection effect. They also suggest that parents who try the public sector are, on average, more inclined to remain if everything goes well (type 1).

4 Empirical Analysis

We divide our empirical section into three subsections. In the first, we introduce the identification strategy for our causal estimation of the effect of distance, along with the framework for our exploratory analysis of the determinants of acceptance and permanence in public education. The second and third subsections present the main results on the role of distance and the other determinants of parental choices, respectively.

4.1 Identification Strategy

The 2020 ME algorithmic allocation process allows us to exploit the exogenous variation from the randomized assignment of students conditional on the local school offering to recover the causal estimates of the parents' preferences for distance when choosing a school. The randomization process is part of the algorithmic mechanism of Deferred Acceptance (DA), which uses imputed preferences based on the distance of the school to the student's home or her preferred address (ranking the closest available school first and furthest school last), for the closest schools within a 5 kilometer radius.

Conditional on students' preferences for proximity, students are randomly and temporarily allocated to public schools to maximize the likelihood of getting a spot at their most preferred school (i.e., that closest to their home). Therefore, students with similar "preferences," in this case, for geographical proximity, ultimately receive their final allocations by chance in the event that the school is oversubscribed. This randomization allows us to compare their behavior. If there are sufficient vacancies for all students applying to a particular school by grade and level, no randomization will take place as all students will receive their preferred placement.[17]

Our identification strategy relies on the exogeneity of the assignment of schools between students conditional on these students having similar "preferences" (list of closest schools). A possible threat to this assumption is that parents strategically list their addresses close to their preferred school. However, as can be observed in Figure 6, the likelihood of accepting an assignment is roughly constant at around 80% at every ranking level. If parents were being strategic, we would observe a right-skewed distribution of acceptance where lower-ranked schools have lower rates of acceptance than higher-ranked ones.

4.1.1 The Role of Distance

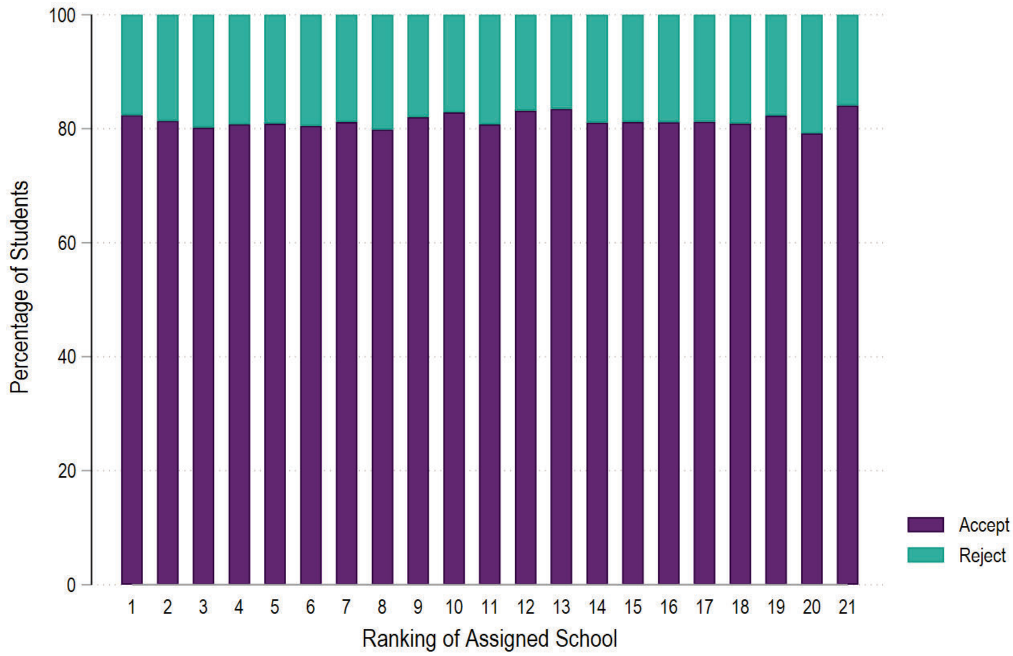
In this section, we introduce our main specification for estimating the causal effect of distance on parents' decisions to accept or reject their assigned school.

$$y_{idsgl} = \alpha_{dgl} + \gamma_s + \delta_i + \beta_1 \cdot \text{Dist}_{is} + \beta_2 \cdot (\text{Cost School Origin})_i + e_{idsgl} \qquad (4)$$

Where y_{idsgl} takes two values: first, Accept_{idsgl} is an indicator variable for student i in district d, assigned school s, grade g and level l, which takes a value of 100 if the student decides to accept her assignment in the first stage and a value of 0 if the student rejects her assignment; and, second, Staying_{idsgl}, which takes a value of 100 if the student decides to stay at her assigned school in 2021 and a value of 0 if the student moves to another school conditional on having accepted the assignment in the previous year. The independent

[17] This is the reasoning for dropping all districts by grade and level where students are all assigned to their first preference

Figure 6: Distribution of Acceptances by Preference Ranking



(a) Note: The ranking of preferences is constructed such that the first ranked school is the closest school to the student.

variables used include Dist_{is}, which is the distance between the assigned school s and student i's home. We include the cost of the school of origin to control for an approximate measure of income level and a dummy to control for the student's gender, denoted by δ_i. We include district d, grade g, and level l fixed effects denoted by α_{dgl} to control for the systematic differences in the feasible set of schools[18], as well as an assigned school level fixed effect, γ_s, which controls for the school's time-invariant characteristics. Finally, e_{idsgl} represents the error term.

[18]This fixed effect imposes that we only exploit variation between students who live in the same district, apply for the same educational level (nursery, primary, or secondary), and grade (1,2,3, etc. depending on the educational level). For more information please refer to table 9

4.1.2 Non-Linear Effects of Distance

Next, we discuss the non-linear effects of distance on the likelihood of a student accepting and staying at the assigned school:

$$y_{idsgl} = \alpha_{dgl} + \gamma_s + \delta_i + \sum_{r=2}^{5} \eta_r \cdot I\{Dist_{is} \in Q_r\} + \beta_2 \cdot (\text{Cost School Origin})_i + e_{idsgl} \quad (5)$$

As in equation (5), we relax the linear constraint on modeling the effect of distance. To do this, we employ specification (4), with a modification on the distance variable. Instead of using the continuous distance variable, we replace it with an indicator that categorizes distances into five discrete values, each representing a 1 km distance range. These categories are [0, 1), [1, 2), [2, 3), [3, 4), and [4, 5). We label each distance group in the distribution as Q_r, where r = 1 denotes the group with the shortest distances, and r = 5 represents the group with the longest distances. Notably, we exclude the smallest categorization (r=1) from our estimation. This exclusion allows us to make comparisons with the group characterized by the shortest distances.

Similarly, in Appendix A.2.4, we employ an alternative categorization of distance. We divide distance into quartiles based on the distribution of home-to-school distances, with the following ranges: [0, 0.86), [0.86, 1.9), [1.9, 3.25), and [3.25, 5].

4.1.3 The Determinants of Parents' Preferences

Next, we introduce the specifications for estimating the determinants of parental preferences in their decision to accept the assigned school and to stay enrolled after a year. The estimates produced by specification (6) are not meant to be interpreted causally, as we cannot disentangle possible correlations between unobservable and observable characteristics of the school, with the possible exception of distance, as described above. The first-stage specification aims to estimate the correlations in parents' preferences for relative school characteristics with respect to their decision to accept or reject their assignment offering. Then, conditional on accepting their assignment in the first stage, we examine the influence of observable characteristics on their decision to stay or leave the assigned school.[19] In particular, we run the following regression:

$$y_{idsgl} = \alpha_{dgl} + \gamma_s + \delta_i + \psi \cdot \text{Dist}_{is} + \mathcal{D}'_{is}\Lambda + \beta_2 \cdot (\text{Cost School Origin})_i + e_{idsgl} \quad (6)$$

Where y_{isdgl} takes two values: first, Accept_{idsgl} is an indicator variable for student i in district d, assigned to school s, level l, and grade g, which takes a value of 100 if the student decides to accept her assignment in the first stage and a value of 0 if the student rejects her assignment; and, second, Staying_{idsgl} which takes a value of 100 if the student decides to stay at her assigned school and a value of 0 if the student moves to another school conditional on having accepted the assignment. The independent variables used include Dist_{is}, which is the distance between the assigned school s and student i's home, and $\mathcal{D}_{is}$, which captures differences in the observable characteristics of the assigned schools s and student i's previous school. To compute the differences, we use the information on school class sizes as measured by the student-to-teacher ratio, average parental education, a measure of academic performance[20], and school size. We also include characteristics that indicate the relevance of teacher quality, such as the proportion of novice teachers (less than five years of experience). We include the cost of the school of origin to control for an approximate measure of students' income level and a dummy to control for the student's gender denoted by δ_i. District d by grade g and level l fixed effects are also included, denoted by α_{dgl}, to control for the systematic differences in the feasible set of schools, as well as an assigned school-level fixed effect, γ_s, which controls for the time-invariant characteristics of the school. Finally, e_{idsgl} represents the error term.

In the next section, we explore specification (4) for estimating β_1. The subsequent section then turns to specification (6), where we estimate ψ and Λ for the first stage and assess the variation in these estimates in the second stage. Finally, we test the robustness of our results by applying more granular fixed effects, where we only compare students

[19] As detailed in the theoretical model, we acknowledge that selection plays a crucial role in the results of this second stage. Therefore, our claim of causality holds primarily for the first stage.

[20] We use average standardized math scores from the ECE.

who have the same top two schools listed in their set of preferences (based on proximity). Comparing students whose top-ranked schools are both identical and oversubscribed closely mimics the randomization mechanism used by the algorithm.[21] The clustering of standard errors is done at the district level, further broken down by grade and level for all regressions shown.

4.2 Main Results

4.2.1 The role of distance

Table 4 presents the results of specification (4), with columns (1) and (3) showing estimates for whether parents accept or reject the assignment, as indicated by the dependent variable. Meanwhile, columns (2) and (4) indicate estimations in which the dependent variable reflects whether parents choose to stay or leave the assigned school, conditional on having accepted the assignment in the previous year. Columns (1) and (2) rely on district, grade, and level fixed effects, while columns (3) and (4) employ a unique fixed effect per set of students with the same top 2 schools. The latter set of fixed effects imposes a greater restriction on the estimation, as only students living very close to each other will have the same top two schools in their preference list. This restriction allows us to directly mimic the randomization devised used by the DA algorithm. As mentioned, specification (4) allows us to causally estimate the effect of distance on the first decision: accept the public school offered or reject it for another alternative (public or private).

The data presented in Table 4 shows the crucial importance of school proximity in parental decisions. Specifically, columns (1) and (3) reveal that for every additional kilometer a child must travel from home to school, the likelihood of parents accepting the assigned school decreases by between 1.5 and 2.03 percentage points (pp). This suggests that parents are attuned to and concerned about the length of their children's daily commute to and from school.

Furthermore, the negative correlation with home-to-school distance is even more pro-

[21] We also refine our fixed effects at a more granular level, in order to compare among those who have the same top 3 schools. These results can be found in Tables 11-15 in the appendix.

nounced in the second stage, with an additional kilometer resulting in an increase of 1.99 to 3.65 pp in the student's likelihood of remaining at the assigned school. This heightened sensitivity suggests that parents' apprehensions about longer distances may become stronger as they progress through the decision-making process.

Recall that all classes remained virtual at the time of these decisions, suggesting that parents' views on distance to school were taken from a long-term perspective. Indeed, in a virtual setting, distance is not a crucial factor. The fact that it still played a role implies that parents may see transitioning to public education as a long-term commitment rather than a short-term solution.

Table 4: The Effect of Distance on the Likelihood of Accepting and Staying at the Assigned School

VARIABLES	(1) Accept	(2) Stay	(3) Accept	(4) Stay
Distance (km)	-1.502***	-1.995***	-2.029***	-3.649***
	(0.125)	(0.245)	(0.508)	(0.626)
Cost School of Origin	-0.014***	-0.019***	-0.015***	-0.026***
	(0.002)	(0.005)	(0.005)	(0.010)
Constant	88.010***	82.920***	83.500***	84.771***
	(0.530)	(1.153)	(1.501)	(3.053)
Observations	74,689	45,671	12,698	7,966
R-squared	0.108	0.174	0.239	0.301
FE: Assigned School	X	X	X	X
Inc. Female	X	X	X	X
FE: Dist-Gr-Lev	X	X		
FE: Same Top 2			X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(a) Note: Standard errors have been clustered at the same level as the randomization takes place (district by educational level and grade.)

In Table 5, we regress the assigned distance to school on observable characteristics of the assigned school. The same fixed effect specifications from Equations 4 and 6 are retained, except for that related to the assigned school level. The lack of systematic correlations between distance and observable characteristics of the assigned school can be interpreted as plausible evidence that distance is also likely to be uncorrelated with unobservable characteristics of the assigned schools. Only the proportion of novice teachers shows a statistically significant relationship with distance. However, the covariates do not exhibit joint significance in their relationship with distance.

Table 5: Relationship between Distance and Covariates

VARIABLES	(1) Distance *First Stage*	(2) Distance *Second Stage*	(3) Distance *First Stage*	(4) Distance *Second Stage*
Mean Math sc (s)	0.133	0.113	0.189	0.453
	(0.168)	(0.167)	(0.318)	(0.335)
Mean Lang sc (s)	-0.156	-0.138	-0.426	-0.855*
	(0.215)	(0.217)	(0.405)	(0.461)
Mean Parent Edu.	-0.046	-0.024	0.032	0.067
	(0.058)	(0.067)	(0.108)	(0.135)
Log(Num Students)	-0.474	-0.402	-0.499	-0.982
	(0.383)	(0.395)	(0.787)	(0.838)
Log(Num Students)^2	0.029	0.023	0.022	0.065
	(0.032)	(0.033)	(0.065)	(0.070)
STR	0.002	0.001	-0.016	-0.017
	(0.006)	(0.007)	(0.013)	(0.015)
Prop. Novice Teachers	0.278**	0.279**	0.245	0.372**
	(0.113)	(0.129)	(0.158)	(0.151)
Cost School of Origin	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	4.104***	3.774***	4.517*	5.639**
	(1.151)	(1.193)	(2.336)	(2.534)
Observations	59,525	39,432	11,448	7,602
R-squared	0.346	0.332	0.395	0.378
Inc. Female	X	X	X	X
FE: Dist-Gr-Lev	X	X		
FE: Same Top 2			X	X
F-test:	8.17	7.52	5.79	5.33
Prob > F:	0.0000	0.0000	0.0000	0.0000

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(a) Note: Standard errors have been clustered at the same level as the randomization takes place (district by educational level and grade.)

4.2.2 Non-Linear Effects: Distance

Table 6 presents the results of specification (5) and Figure 7 displays the estimates for η_r for $r \in \{2, 3, 4, 5\}$ from columns (1) and (2). Mirroring the format of the previous table, columns (1) and (3) present the results of the first stage, while columns (2) and (4) present the results of second stage. In all specifications, we exclude the smallest distance category ($r = 1$). Therefore, the coefficients are meant to be interpreted as the differences between each group relative to the first category as a reference group.

Similarly, in Appendix A.2.4 we show the results for another categorization, where the groups are separated by quartiles.

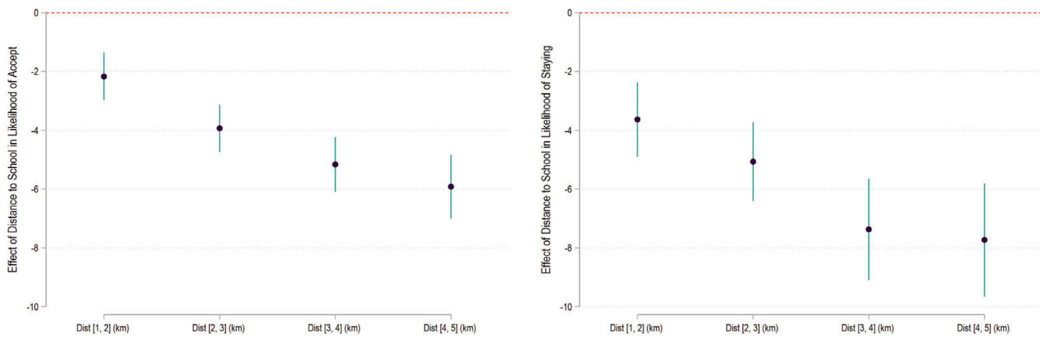
Table 6: Non-Linear Effects of Distance by Sections

VARIABLES	(1) Accept	(2) Stay	(3) Accept	(4) Stay
Dist [1, 2] (km)	-2.170***	-3.634***	-1.884	-7.065***
	(0.412)	(0.643)	(1.294)	(1.985)
Dist [2, 3] (km)	-3.933***	-5.068***	-5.776***	-7.165***
	(0.409)	(0.681)	(1.475)	(1.912)
Dist [3, 4] (km)	-5.161***	-7.370***	-6.122***	-11.949***
	(0.473)	(0.874)	(1.871)	(2.658)
Dist [4, 5] (km)	-5.919***	-7.729***	-7.189***	-14.479***
	(0.550)	(0.976)	(2.167)	(2.750)
Cost School of Origin	-0.014***	-0.019***	-0.015***	-0.026***
	(0.002)	(0.005)	(0.005)	(0.010)
Constant	87.681***	82.689***	82.649***	83.828***
	(0.570)	(1.107)	(1.450)	(3.000)
Observations	74,675	45,659	12,697	7,965
R-squared	0.108	0.174	0.239	0.301
FE: Assigned School	X	X	X	X
Inc. Female	X	X	X	X
FE: Dist-Gr-Lev	X	X		
FE: Same Top 2			X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 7 (both panels) shows a linear relationship between the home-to-school distance and the likelihood of accepting and staying at the assigned school when allowing for non-linear effects in each distance category. We find that the coefficients of the effect on the likelihood of accepting the assigned school, relative to the smallest distance category, vary between -2.17 pp and -5.92 pp. We can only distinguish the coefficients in the [1,2) km range from those in the [3,4) and [4,5) km ranges. The impact of distance on the probability of students accepting the assigned school almost triples in magnitude from -2.17 in the [1,2)

Figure 7: Non-Linear Effects of Distance by Sections



(a) On the likelihood of accepting

(b) On the likelihood of staying

km range to -5.92 in the [4,5) km range. The coefficients of the effect on the likelihood of staying at the assigned school, relative to the smallest distance category, vary between -3.63 pp and -7.73 pp (these effects can reach up to -14.48 pp when considering different sets of fixed effects).

When students attend a school located more than 3 km away from their family's home, there is a 7.37 pp decrease in the likelihood of them remaining at that assigned school relative to students who attend a school that is within 1 km of their family's home. This indicates that proximity is a significant factor for families when assessing the attractiveness of a school.

In Figure 7, panel (a) continues to display a linear trend in the coefficients. However, in panel (b), the point estimates of the coefficients appear to stabilize after 3km. We can only distinguish the coefficients in the [1,2) km range from those in the [3,4) and [4,5) km ranges. The impact of distance on the probability of students remaining at the assigned school more than doubles in magnitude from -3.63 in the [1,2) km range to -7.73 in the [4,5) km range. As detailed in the main text, parents' decisions in the second stage appear to be highly influenced by the school's distance.

4.2.3 The Determinants of Parents' Preferences

Table 7 presents the results of specification (6). As with the previous tables, columns (1) and (3) display the results of the first stage, while columns (2) and (4) present the results of the second stage. However, due to the presence of unobserved variables that may be correlated with our observed variables, specification (6) does not allow us to infer causality. Consequently, we can only estimate correlations between the attributes of the schools that students are assigned to and the characteristics that families experienced in their previous private school, in relation to the likelihood of accepting and staying at the given assignment.

Specifically, the table considers the relative characteristics of the public schools to which students are assigned compared to the characteristics of the private schools they previously attended. As above, columns (1) and (2) rely on district, grade, and level fixed effects and columns (3) and (4) use the top two schools fixed effects. It is noteworthy that there is a significant decrease in the number of observations between Table 4 and column (1) of Table 7. The discrepancy is due to the exclusion of clusters of observations that lack adequate variation in the acceptance rate at our chosen observation level (district by grade and level). Additionally, some of the covariates in our analysis have missing values for some students.

As seen in the previous section, the estimated effect of distance remains consistent with the results in Table 4, highlighting the negative effect of distance on both the likelihood of accepting an assignment as well as remaining in the assigned schools. Beyond the role of distance, our data suggests that parents also care about school quality, which can be assessed through scores on standardized math tests. We use the school's average math score on the ECE exams as an indicator of its quality, in line with other papers in the literature (Hofflinger et al. (2020), Burgess et al. (2015), Hastings et al. (2006)). In particular, we observe that an increase of one standard deviation in the difference in mean math scores between the assigned and origin schools is correlated with an increase of 5.35 to 7.66 pp in

Table 7: The Relationship of School Characteristics to the Likelihood of Accepting and Staying at the Assigned School

VARIABLES	(1) Accept	(2) Stay	(3) Accept	(4) Stay
Distance (km)	-1.227***	-1.882***	-1.732***	-3.395***
	(0.139)	(0.228)	(0.593)	(0.568)
Dif. Mean Math sc (s)	5.351***	4.301***	7.662***	6.476**
	(0.631)	(0.795)	(1.792)	(2.533)
Dif. Mean Parent Edu.	1.824***	3.827***	2.388***	6.093***
	(0.370)	(0.424)	(0.844)	(1.549)
Log(Ratio Num St.)	2.318***	4.180***	1.772	6.233***
	(0.577)	(0.731)	(2.104)	(1.678)
Log(Ratio Num St.)^2	-0.039	-0.568***	0.162	-1.204**
	(0.162)	(0.176)	(0.562)	(0.548)
Dif. STR	-0.098***	-0.050	-0.042	-0.075
	(0.034)	(0.046)	(0.093)	(0.126)
Dif. Prop. Novice Teachers	-1.555**	-1.379**	-1.229	-0.462
	(0.691)	(0.660)	(1.910)	(2.351)
Cost School of Origin	-0.005***	-0.007***	-0.002	-0.003
	(0.001)	(0.003)	(0.003)	(0.009)
Constant	83.439***	82.276***	78.330***	81.749***
	(0.793)	(1.080)	(2.537)	(3.297)
Observations	51,484	33,419	9,476	5,986
R-squared	0.116	0.175	0.263	0.337
FE: Assigned School	X	X	X	X
Inc. Female	X	X	X	X
FE: Dist-Gr-Lev	X	X		
FE: Same Top 2			X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(a) Note: Standard errors have been clustered at the same level as the randomization takes place (district by educational level and grade.)

the likelihood of accepting the assignment and an increase of 4.3 to 6.48 pp in the likelihood of staying at the assigned school. The consistent student performance observed in both the initial and subsequent stages underlines the importance parents place on school quality. This result aligns with findings reported by Beuermann et al. (2023) and Abdulkadiroğlu et al. (2020), among others, who document that parents prefer higher-quality schools.

In addition to their concern over distance and school quality, parents also value peer demographics. When there is a unit increase in the average level of parent education[22] (denoting an increase of one level on the academic qualification hierarchy detailed earlier) in the assigned school relative to the private school of origin, the likelihood of accepting the assigned school increases by between 1.82 pp and 2.39 pp. This suggests that peer demographics is a relevant consideration in school choice.

[22] As shown in Table 2, the average parental education level is 4.9, just under "complete secondary education."

Furthermore, point estimates of the inclination to remain at the assigned school increase by 3.83 to 6.09 pp for an increase in average parent education level. As previously discussed, parents transferring out of the private system typically have higher education levels than their peers in the assigned public schools. This discrepancy highlights parents' preference for maintaining positive peer demographics, indicating a desire for schools where the parental education level is consistent with their prior experience.

Moreover, parental education levels can serve as an indirect measure of other valued school features that are not strictly related to peer demographics. For instance, parents might assume that schools with more educated parents are higher performing (possibly due to increased advocacy and pressure from these parents), safer, or have other valued aspects that are not directly related to student performance. In the absence of test score data, peer demographics might also be used by parents as a proxy for student achievement.

Relative school size is another factor that parents consider. A unit increase in the log ratio in school size - measured by the number of enrolled students - between the assigned and origin school is associated with an greater probability of accepting and staying at the assigned school. Parents could view the size of the assigned school as an indicator of its quality or popularity. However, the relationship is not linear.[23] The likelihood of acceptance and retention is notably higher when the assigned school's size greatly exceeds that of the private school of origin.

The results furthermore show that in the assignment stage, parents are less likely to accept being assigned to schools with higher student-teacher ratios. However, this result does not hold if we consider our more stringent set of fixed effects (see column (3)). Interestingly, this variable does not have a significant effect on the decision to remain in the assigned school. We also observe that parents are less likely to accept and stay in assigned schools with a higher proportion of novice teachers, another dimension of teacher quality. Although this result does not hold in our more stringent fixed effect specification,

[23] We test for joint significance of the quadratic terms: in the first column, the F-statistic is 14.03, while in the second column, it is 17.55, i.e., both are jointly significant.

it suggests that families prefer schools with more experienced teachers.

Finally, we observe that families who pay higher tuition at their original private schools are less likely to accept an assignment to a public school. Cost of the school can be seen as an indicator of a family's purchasing power. Those with more financial resources, who thus have access to a broader range of schools, may be less inclined to switch to the public sector. Alternatively, parents who can afford higher private school fees may equate higher costs with superior quality. They might consequently be less willing to transition their children away from what they perceive to be higher-quality private institutions compared to public schools. However, the significance of this coefficient drops considering our more stringent set of fixed effects.

Our findings are consistent with evidence reported by Jacob and Lefgren (2007) and Burgess et al. (2015), which shows that the families prefer schools that are higher-performing, with better peer demographics, and located closer to their residence. Using data from a centralized matching system after eliminating school attendance zones in New Orleans post–Hurricane Katrina, Harris and Larsen (2017) find that parents prefer schools with higher test scores and higher added value, and that they also weigh non-academic factors. These factors include the availability of after-school care, extracurricular activities, whether the child had a sibling in the school, and the school's proximity to home. Additionally, (Glazerman and Dotter, 2017) report that commuting distance, peer demographics, and academic indicators play essential roles in school choice. Their study also highlights considerable heterogeneity in preferences across parent demographics.

In Table 8, we standardize the same variables detailed in Table 7, allowing us to discern the relative significance of each dimension in influencing decisions to accept and remain at the assigned schools. Notably, we observe that many factors impacting the decision to accept assigned schools have a similar weight.

During the initial decision-making phase, distance to the school, school size and school quality (proxied by standardized math test scores) appear to be the most important considerations (closely followed by parental education). These are also the factors that parents are most likely to be aware of before enrollment. However, once families have had direct experience with the school, other factors such as the variations in the average education level of other parents become more relevant in their continued commitment to stay at the school. For instance, when the variables are standardized, the effect of the difference in standardized math test scores decreases from 3 pp in the first stage to 2.41 in the second stage, while the effect of the difference in the mean of parental education increases from 1.93 pp. in the first stage to 4.04 in the second.

Table 8: The Relationship of School Characteristics to the Likelihood of Accepting and Staying at the Assigned School - Standardized

VARIABLES	(1) Accept	(2) Stay	(3) Accept	(4) Stay
Distance (km)	-2.453***	-3.764***	-3.463***	-6.789***
	(0.277)	(0.456)	(1.186)	(1.135)
Dif. Mean Math sc (s)	3.002***	2.413***	4.298***	3.633**
	(0.354)	(0.446)	(1.005)	(1.421)
Dif. Mean Parent Edu.	1.925***	4.039***	2.520***	6.431***
	(0.390)	(0.447)	(0.890)	(1.635)
Log(Ratio Num St.)	2.618***	4.722***	2.001	7.041***
	(0.652)	(0.825)	(2.376)	(1.895)
Log(Ratio Num St.)^2	-0.117	-1.681***	0.480	-3.565**
	(0.480)	(0.523)	(1.666)	(1.622)
Dif. STR	-0.920***	-0.470	-0.398	-0.707
	(0.321)	(0.435)	(0.878)	(1.194)
Dif. Prop. Novice Teachers	-0.587**	-0.520**	-0.463	-0.174
	(0.261)	(0.249)	(0.720)	(0.887)
Cost School of Origin	-0.991***	-1.443***	-0.344	-0.541
	(0.262)	(0.509)	(0.665)	(1.734)
Constant	80.045***	75.131***	73.951***	69.688***
	(0.083)	(0.164)	(0.391)	(0.574)
Observations	51,484	33,419	9,476	5,986
R-squared	0.116	0.175	0.263	0.337
FE: Assigned School	X	X	X	X
Inc. Female	X	X	X	X
FE: Dist-Gr-Lev	X	X		
FE: Same Top 2			X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(a) Note: Standard errors have been clustered at the same level as the randomization takes place (district by educational level and grade.)

In contrast to specification (6), we could have used the observable characteristics of the assigned schools when evaluating the decisions to accept and remain enrolled. Table 17 in the appendix presents the results of these specifications. While there are clear similarities between these results and those presented in Table 7, particularly concerning the influence of distance and the significance of peer quality, there are also some interesting differences. Most prominently, the quality of the assigned school does not appear to influence the decision to accept the school assignment or the decision to continue attending. Interestingly, the school's size only becomes a significant factor in the second stage, when deciding whether to stay at the assigned school.[24] We also observe that the proportion of novice teachers appears relevant and robust in the acceptance stage but only after using our more stringent set of fixed-effects. Younger teachers might be viewed favorably because they are perceived as being more adept with technology. Given that education was primarily delivered remotely during this period, technological expertise could be seen as an indicator of quality.

[24]We test the joint significance in the first stage and find null results for both coefficients of school size (p-value = 0.3)

5 Discussion

During the first year of the COVID-19 pandemic, Peru experienced an unprecedented rise in the demand for public schools from students previously enrolled in private schools, with more than 100,000 pupils requesting a transfer. Different factors likely contributed to this surge, including the extended school closures during the first year of the pandemic, the inability of many private schools to switch to online education, and the economic burden of paying private school tuition during an economic crisis.

In response, the Ministry of Education designed a centralized assignment system to integrate these students into the public sector. The inherent experimental variation of this system presents a unique opportunity to better understand the priorities of this specific group of parents: middle- and lower-middle-class families who would otherwise enroll their children in private schools (Balarin et al., 2019). It furthermore offers the possibility of gaining valuable insights useful for governments seeking to strengthen the supply of public schooling and potentially reverse pre-pandemic trends in public and private education.

Our paper sheds light on what this group of parents value in a public school before and after their experience with the assigned public school. We find that parents prioritize proximity, academic quality, and student demographics when deciding to accept and continue in the public school system. When comparing the impact of these features on the decision to accept a public school placement and then, one year on, to remain at the assigned public school, we find that absolute distance to the school and peer demographics play a more significant role than academic quality in the decision to stay. Nevertheless, all three factors remain relevant in both phases of decision-making.

The results of this study have important implications for policy.[25] First, consistent with previous evidence and controlling for other factors, parents care about student demographics. In the context of Latin America and the Caribbean, public schools tend to

[25] Although, care should be taken in extrapolating these results since, in this setting, parents only learned about the public school imperfectly through the online remote learning experience. Had parents experienced the public school in person, the outcomes may have differed.

be less diverse and serve the most disadvantaged students. Therefore, attracting a more diverse clientele into the public sector is essential for fostering diversity and drawing more middle-class families. The segregation of students may otherwise become more severe, and public schools more stigmatized. Second, even after controlling for other relevant factors, the school's academic quality remains a principal concern for parents in the decision to both choose and stay in the assigned school. It is therefore imperative that the Peruvian government continue to focus on reforms that improve the quality of public schools. For instance, over the last decade, the government has instituted reforms to modernize teachers' career paths (Elacqua et al., 2018) and attract and select more effective teachers into the profession.

Finally, a crucial insight concerns the role of experimentation in shaping parents' enrollment decisions. The policy implication is that short-term incentives that provide firsthand experiences with public services, like education, can have long-term consequences for a significant group of parents. Surprisingly, only 15.68%[26] of the students who experienced the assigned public school in 2020 returned to the private sector in 2021. Despite evidence that, all else being equal, private schools do not have higher achievement than public schools (Balarin, 2015), parents in Peru perceive private schools to be better.[27] Publishing objective school-level information on test scores or value-added measures may reduce misconceptions about public and private school quality. The Peruvian government might also run campaigns to increase citizens' awareness of the objective attributes of public schools.

[26] Just 10,181 students out of the 64,916 who experienced at least a few months in their assigned public school in 2020 returned to the private sector after a year.

[27] *Este País*, May 1^{st}, 2014

References

A. Abdulkadiroğlu, N. Agarwal, and P. A. Pathak. The welfare effects of coordinated assignment: Evidence from the new york city high school match. *American Economic Review*, 107(12):3635–89, 2017.

A. Abdulkadiroğlu, P. A. Pathak, J. Schellenberg, and C. R. Walters. Do parents value school effectiveness? *American Economic Review*, 110(5):1502–39, 2020.

B. Abizanda, G. Almeyda, E. Arias Ortiz, C. Berlanga, I. Bornacelly, M. S. Bos, E. Díaz, X. Dueñas, G. Elacqua, A. Elías, R. Fernández-Coto, V. Frisancho, V. A. García Moreno, J. C. Hernández Cardozo, D. Hincapie, J. F. Margitic, L. Marotta, M. M. Mateo-Berganza Díaz, A. Morduchowicz, F. Muñoz, E. Näslund-Hadley, M. Ruiz-Arranz, A. Thailinger, F. J. Valverde Rodríguez, E. Vezza, and P. Zoido. How to reboot education post-pandemic: Delivering on the promise of a better future for youth. *IDB Monograph ; 1012*, 2022.

N. Agarwal and P. Somaini. Demand analysis using strategic reports: An application to a school choice mechanism. *Econometrica*, 86(2):391–444, 2018.

M. Balarin. The default privatization of peruvian education and the rise of low-fee private schools: Better or worse opportunities for the poor? 2015.

M. Balarin, C. Fontdevila, P. Marius, and M. F. Rodríguez. Educating on a budget: The subsistence model of low-fee private schooling in peru. *Education Policy Analysis Archives*, 27(132), 2019.

D. W. Beuermann, C. K. Jackson, L. Navarro-Sola, and F. Pardo. What is a good school, and can parents tell? evidence on the multidimensionality of school output. *The Review of Economic Studies*, 90(1):65–101, 2023.

S. Burgess, E. Greaves, A. Vignoles, and D. Wilson. What parents want: School preferences and school choice. *The Economic Journal*, 125(587):1262–1289, 2015.

G. Elacqua. The impact of school choice and public policy on segregation: Evidence from chile. *International Journal of Educational Development*, 32(3):444–453, 2012.

G. Elacqua, M. Schneider, and J. Buckley. School choice in chile: Is it class or the classroom? *Journal of Policy Analysis and Management*, pages 577–601, 2006.

G. Elacqua, D. Hincapié, E. Vegas, M. Alfonso, V. Montalva, and D. Paredes. *Profesión: Profesor en América Latina:¿ Por qué se perdió el prestigio docente y cómo recuperarlo?* Inter-American Development Bank, 2018.

G. Elacqua, C. Méndez, and M. Navarro. Colegios privados en tiempos de covid-19 en américa latina y el caribe. *IADB Technical Note*, 2022. doi: 10.18235/0004434.

S. Glazerman and D. Dotter. Market signals: Evidence on the determinants and consequences of school choice from a citywide lottery. *Educational Evaluation and Policy Analysis*, 39(4):593–619, 2017.

D. N. Harris and M. F. Larsen. Demand, information, and the market for schooling: Evidence on revealed preferences from post-katrina new orleans. *Education Research Alliance for New Orleans*, 2017.

J. S. Hastings, T. J. Kane, and D. O. Staiger. Parental preferences and school competition: Evidence from a public school choice program, 2005.

J. S. Hastings, T. J. Kane, and D. O. Staiger. Preferences and heterogeneous treatment effects in a public school choice lottery, 2006.

A. Hofflinger, D. Gelber, and S. T. Cañas. School choice and parents' preferences for school attributes in chile. *Economics of Education Review*, 74:101946, 2020.

B. A. Jacob and L. Lefgren. What do parents value in education? an empirical investigation of parents' revealed preferences for teachers. *The Quarterly Journal of Economics*, 122 (4):1603–1637, 2007.

A. Kapor, C. Neilson, and S. Zimmerman. Heterogeneous beliefs and school choice. *American Economic Review*, 2016.

B. Kleitz, G. R. Weiher, K. Tedin, and R. Matland. Choice, charter schools, and household preferences. *Social science quarterly*, pages 846–854, 2000.

F. J. Murillo, G. Belavi, L. M. Pinilla, et al. Public-private school segregation in spain. *Universitat Autonoma de Barcelona Papers*, 2018.

M. Schneider, P. Teske, M. Marshall, and C. Roch. Shopping for schools: In the land of the blind, the one-eyed parent may be enough. *American Journal of Political Science*, pages 769–793, 1998.

A. Singh, A. Park, and S. Dercon. School meals as a safety net: An evaluation of the midday meal scheme in india. *Economic Development and Cultural Change*, 62(2):275–306, 2014. doi: 10.1086/674097. URL https://doi.org/10.1086/674097.

A Appendix

A.1 Theorical Appendix

A.1.1 Proof Lemma 3.0.3

Proof. If we assume that G is a uniform distribution, the derivative of the probability of experimenting with respect to V_{PU} is

$$\frac{\partial \mathbb{P}(try)}{\partial V_{PU}} = g \cdot \left[\mu_0 H_1'(V_{PU}) + (1-\mu_0) H_2'(V_{PU})\right] \tag{7}$$

where $H_i'(V_{PU}) = \bar{\alpha}_i + \int_{\frac{V_{PR}^{(2)}}{V_{PU}}}^{K} \alpha \cdot dF_i(\alpha)$

and the derivative of the probability of staying with respect to V_{PU} is:

$$\begin{aligned}\frac{\partial \mathbb{P}(stay)}{\partial V_{PU}} =& g\left[\mu_1 \int_{\alpha=1}^{K} \alpha dF_1(\alpha) + (1-\mu_1)\int_{\alpha=1}^{K} \alpha dF_2(\alpha)\right] + \\ & \frac{\partial \mu_1}{\partial V_{PU}}\left[\int_{\alpha=1}^{K} G(L(\alpha V_{PU}))dF_1(\alpha) - \int_{\alpha=1}^{K} G(L(\alpha V_{PU}))dF_2(\alpha)\right]\end{aligned} \tag{8}$$

where $\frac{\partial \mu_1}{\partial V_{PU}} = \frac{\mu_0(1-\mu_0)g\cdot\left[H_1'(V_{PU})G(H_2(V_{PU})) - H_2'(V_{PU})G(H_1(V_{PU}))\right]}{(\mu_0 G(H_1(V_{PU})) + (1-\mu_0)G(H_2(V_{PU})))^2}$.

Let $\lambda(V_{PU}) = \frac{\partial \mu_1}{\partial V_{PU}}\left[\int_{\alpha=1}^{K} G(L(\alpha V_{PU}))dF_1(\alpha) - \int_{\alpha=1}^{K} G(L(\alpha V_{PU}))dF_2(\alpha)\right]$. If we substract (8) - (7) we obtain the difference in partial derivatives:

$$\underbrace{(\mu_1-\mu_0)(\bar{\alpha_1}-\bar{\alpha_2}) + \lambda(V_{PU})}_{\text{Selection effect}} - \underbrace{\left[(1-\mu_0)\int_{\alpha=\frac{V_{PR}^{(2)}}{V_{PU}}}^{\alpha=K} \alpha \cdot dF_2(\alpha) + \mu_0 \int_{\alpha=\frac{V_{PR}^{(2)}}{V_{PU}}}^{\alpha=K} \alpha \cdot dF_1(\alpha)\right]}_{\text{Horizon effect}}$$

□

A.2 Empirical Appendix

A.2.1 Levels and grades

Table 9: Students by level and grade

Level/Grade	N
Level: *Initial*	16,243
Level: *Primary*	42,581
Grades:	
1	8,457
2	8,853
3	7,421
4	6,512
5	5,921
6	5,417
Level: *Secondary*	19,728
Grades:	
1	4,328
2	4,078
3	3,808
4	3,926
5	3,588

A.2.2 The Role of Distance - Difference

In this section, we enhance our primary specification (4), which estimates the causal effect of distance on parents' decisions to accept or reject their assigned school. We incorporate the difference in distance between the assigned school and the school of origin. We also investigate its influence on the decision to stay or leave after previously accepting their assignment.

$$y_{idsgl} = \alpha_{dgl} + \gamma_s + \delta_i + \beta_1 \cdot \text{Dist}_{is} + \beta_2 \cdot \text{Dif. Dist}_{is} + \beta_3 \cdot (\text{Cost School Origin})_i + e_{idsgl} \quad (9)$$

Where y_{idsgl} takes two values: first, Accept_{idsgl} is an indicator variable for student i in district d, assigned school s, grade g, and level l, which takes a value of 100 if the student decides to accept her assignment in the first stage and a value of 0 if the student rejects her assignment; and, second, Staying_{idsgl} which takes a value of 100 if the student decides to stay at her assigned school in 2021 and a value of 0 if the student moves to another school, conditional on having first accepted the assignment. The independent variables used include Dist_{is}, which is the distance between the assigned school s and student i's home, and Dif. Dist_{is}, which is the difference between the distances from the student i's home and assigned school s compared to the distance between the school of origin and student i's home. We include the cost of the school of origin as an approximate measure of income level and a dummy to control for the student's gender, denoted by δ_i. We include district d by grade g and level l fixed effects, denoted by α_{dgl}, to control for the systematic differences in the feasible set of schools, as well as an assigned school level fixed effect, γ_s, which controls for the time-invariant characteristics of the school. Finally, e_{idsgl} represents the error term.

The results are as follows:

Table 10: The Effect of Distance on the Likelihood of Accepting and Staying at the Assigned School

VARIABLES	(1) Accept	(2) Stay	(3) Accept	(4) Stay
Distance (km)	-1.173***	-2.082***	-1.311***	-3.837***
	(0.135)	(0.247)	(0.451)	(0.612)
Dif. Distance (km)	-0.345***	0.100*	-0.714***	0.183
	(0.041)	(0.059)	(0.126)	(0.153)
Cost School of Origin	-0.014***	-0.019***	-0.015***	-0.025***
	(0.002)	(0.005)	(0.005)	(0.010)
Constant	87.285***	83.061***	82.122***	84.958***
	(0.531)	(1.126)	(1.475)	(2.970)
Observations	73,455	44,971	12,543	7,872
R-squared	0.109	0.173	0.243	0.303
FE: Assigned School	X	X	X	X
Inc. Female	X	X	X	X
FE: Dist-Gr-Lev	X	X		
FE: Same Top 2			X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(a) Note: Standard errors have been clustered at the same level as the randomization takes place (district by educational level and grade.)

Interestingly, while the distance, in absolute terms, impacts the likelihood of accepting the assignment, the relative distance between the assigned school and the school of origin (Dif. Distance) also negatively impacts the chance of accepting the assignment. In essence, families are unwilling to accept school assignments farther away from their homes than their school of origin. This comparative framing is relevant at the first stage, but not when deciding to stay at the newly assigned school, as depicted in columns (2) and (4), where the signs are positive and not statistically significant (at 5%). This observation may suggest that the reference school changes between stages, implying that in the decision to stay or leave, the previously attended school becomes less relevant in the decision-making framework of the parents.

Furthermore, to illustrate the potential independence between the assigned distance to school and other observable characteristics of the assigned school, we examine Table 11. In this table, the dependent variable is the assigned distance to school, and the covariates are differences. The same fixed effect specifications from Equations 4 and 6 are retained. The absence of systematic correlations between distance and observable relative characteristics of the assigned school can be interpreted as plausible evidence that distance would also likely be uncorrelated with unobservable characteristics of the assigned schools.

Table 11: Relation Distance and Difference in Covariates

VARIABLES	(1) Distance *First Stage*	(2) Distance *Second Stage*	(3) Distance *First Stage*	(4) Distance *Second Stage*
Dif. Mean Math sc (s)	-0.009	-0.034	-0.013	0.111
	(0.040)	(0.049)	(0.073)	(0.076)
Dif Mean Lang sc (s)	0.013	0.030	0.028	-0.097
	(0.040)	(0.049)	(0.071)	(0.069)
Dif. Mean Parent Edu.	0.028	0.029	-0.027	-0.066**
	(0.018)	(0.019)	(0.021)	(0.030)
Log(Ratio Num St.)	0.021	0.037*	0.015	0.043
	(0.016)	(0.020)	(0.030)	(0.049)
Log(Ratio Num St.)^2	-0.000	-0.006	-0.002	-0.012
	(0.005)	(0.005)	(0.010)	(0.012)
Dif. STR	-0.003***	-0.003**	0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.003)
Dif. Prop. Novice Teachers	-0.006	-0.022	-0.007	-0.082*
	(0.019)	(0.022)	(0.032)	(0.045)
Cost School of Origin	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	2.161***	2.132***	2.030***	1.929***
	(0.033)	(0.036)	(0.049)	(0.057)
Observations	51,484	33,419	9,476	5,986
R-squared	0.455	0.447	0.620	0.630
FE: Assigned School	X	X	X	X
Inc. Female	X	X	X	X
FE: Dist-Gr-Lev	X	X		
FE: Same Top 2			X	X
F-test:	2.24	2.04	2.42	2.22
Prob > F:	0.0000	0.0000	0.0000	0.0000

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

(a) Note: Standard errors have been clustered at the same level as the randomization takes place (district by educational level and grade.)

A.2.3 The Role of Distance - Robustness Check

Tables 12 and 13 present a robustness check for specifications 4 and 5, respectively. We have added columns (5) and (6) to Tables 4 and 10. For this analysis, we employ the same top 3 schools as fixed effects, which imposes a stricter requirement for household proximity.

Upon incorporating these more stringent fixed effects, we observe an increase in the magnitude of the effect of distance. Notably, there is a significant reduction in the number of observations from Figure 1 to column (1), which is attributed to the smaller group sizes when employing the same top two and same top three fixed effects, as there is no excess demand in most of these groups.

Table 12: The Effect of Distance on the Likelihood of Accepting and Staying at the Assigned School - Robustness Check

VARIABLES	(1) Accept	(2) Stay	(3) Accept	(4) Stay	(5) Accept	(6) Stay
Distance (km)	-1.502***	-1.995***	-2.029***	-3.649***	-2.954***	-4.323***
	(0.125)	(0.245)	(0.508)	(0.626)	(0.582)	(0.991)
Cost School of Origin	-0.014***	-0.019***	-0.015***	-0.026***	-0.023***	-0.040***
	(0.002)	(0.005)	(0.005)	(0.010)	(0.007)	(0.015)
Constant	88.010***	82.920***	83.500***	84.771***	84.403***	86.919***
	(0.530)	(1.153)	(1.501)	(3.053)	(2.298)	(4.513)
Observations	74,689	45,671	12,698	7,966	7,204	4,438
R-squared	0.108	0.174	0.239	0.301	0.292	0.338
Inc. Female	X	X	X	X	X	X
FE: Dist-Gr-Lev	X	X				
FE: Same Top 2			X	X		
FE: Same Top 3					X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 13: The Effect of Distance on the Likelihood of Accepting and Staying at the Assigned School - Robustness Check

VARIABLES	(1) Accept	(2) Stay	(3) Accept	(4) Stay	(5) Accept	(6) Stay
Distance (km)	-1.173***	-2.082***	-1.311***	-3.837***	-2.124***	-4.428***
	(0.135)	(0.247)	(0.451)	(0.612)	(0.518)	(0.998)
Dif. Distance (km)	-0.345***	0.100*	-0.714***	0.183	-0.900***	0.154
	(0.041)	(0.059)	(0.126)	(0.153)	(0.206)	(0.214)
Cost School of Origin	-0.014***	-0.019***	-0.015***	-0.025***	-0.025***	-0.039***
	(0.002)	(0.005)	(0.005)	(0.010)	(0.008)	(0.014)
Constant	87.285***	83.061***	82.122***	84.958***	82.910***	87.082***
	(0.531)	(1.126)	(1.475)	(2.970)	(2.365)	(4.451)
Observations	73,455	44,971	12,543	7,872	7,121	4,370
R-squared	0.109	0.173	0.243	0.303	0.295	0.340
FE: Assigned School	X	X	X	X	X	X
Inc. Female	X	X	X	X	X	X
FE: Dist-Gr-Lev	X	X				
FE: Same Top 2			X	X		
FE: Same Top 3					X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

A.2.4 Non-Linear Effect: Distance - Quartiles

In this section, we adopt an alternative categorization of distance, as referenced in 4.1.2. We partition distance into quartiles based on the distribution of distance to school. Consequently, these groups cover the following ranges: [0, 0.86), [0.86, 1.9), [1.9, 3.25), and [3.25, 5].

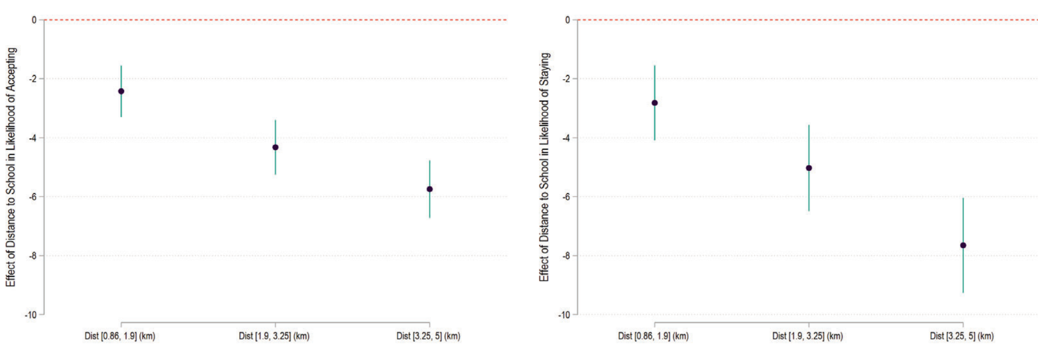
Figure 8 (both panels) shows a surprisingly linear relationship between the distance to the school and the likelihood of accepting and staying at the assigned school when allowing for non-linear effects in each distance category. We find that the coefficients of the effect on the likelihood of accepting the assigned school for each group, when compared to the smallest distance quartile, vary between -2.42 pp up to -5.73 pp. Also, the coefficients of the effect on the likelihood of staying at the assigned school for each group, when compared to the smallest quartile of distance, vary between -2.82 pp and -7.65 pp (the effects can reach up to -13.28 pp when considering different sets of fixed effects). In addition we observe that the size of the effect of the largest quartile group is statistically different from that of the second quartile ([0.79,1.8)) for both the decisions to accept and stay at the assigned school.

Table 14: Non-Linear Effects of Distance by Quartiles

VARIABLES	(1) Accept	(2) Stay	(3) Accept	(4) Stay
Dist. [0.86, 1.9] (km)	-2.424***	-2.819***	-2.170*	-5.391**
	(0.444)	(0.646)	(1.283)	(2.406)
Dist. [1.9, 3.25] (km)	-4.325***	-5.028***	-6.536***	-7.502***
	(0.472)	(0.740)	(1.568)	(2.078)
Dist. [3.25, 5] (km)	-5.746***	-7.650***	-6.252***	-13.281***
	(0.495)	(0.818)	(1.783)	(2.498)
Cost School of Origin	-0.014***	-0.019***	-0.014***	-0.026***
	(0.002)	(0.005)	(0.005)	(0.010)
Constant	88.059***	82.700***	83.039***	83.809***
	(0.601)	(1.044)	(1.587)	(2.925)
Observations	74,689	45,671	12,698	7,966
R-squared	0.108	0.174	0.239	0.301
FE: Assigned School	X	X	X	X
Inc. Female	X	X	X	X
FE: Dist-Gr-Lev	X	X		
FE: Same Top 2			X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 8: Non-Linear Effects of Distance by Quartiles



(a) On likelihood of accepting

(b) On likelihood of staying

A.2.5 The Determinants of Parents' preferences - Robustness Check

In this section, we conduct a robustness check of specification 6. The results are presented in Table 15, and the results with the standardized variables are shown in Table 16. The first four columns are the same as in Table 7. Columns (5) and (6) change the set of fixed effects. In these columns, we use a unique fixed effect per set of the same top three schools, which means an even stricter imposition on the closeness between households.

After applying the more stringent fixed effects, we do not observe significant differences in the second-stage specification compared to the one with the same top two fixed effect. However, there is a change in both magnitude and statistical significance in the first stage. Notably, there is a drop in the number of observations between Figure 1 and Table 15, specifically from Figure 1 to column (5).

Table 15: The Effect of School Characteristics on the Likelihood of Accepting and Staying at the Assigned School - Robustness Check

VARIABLES	(1) Accept	(2) Stay	(3) Accept	(4) Stay	(5) Accept	(6) Stay
Distance (km)	-1.227***	-1.882***	-1.732***	-3.395***	-2.961***	-3.524***
	(0.139)	(0.228)	(0.593)	(0.568)	(0.764)	(0.842)
Dif. Mean Math sc (s)	5.351***	4.301***	7.662***	6.476**	10.441***	9.101**
	(0.631)	(0.795)	(1.792)	(2.533)	(2.627)	(4.280)
Dif. Mean Parent Edu.	1.824***	3.827***	2.388***	6.093***	0.656	6.599***
	(0.370)	(0.424)	(0.844)	(1.549)	(1.543)	(2.131)
Log(Ratio Num St.)	2.318***	4.180***	1.772	6.233***	2.867	5.852**
	(0.577)	(0.731)	(2.104)	(1.678)	(3.094)	(2.244)
Log(Ratio Num St.)^2	-0.039	-0.568***	0.162	-1.204**	0.261	-1.137
	(0.162)	(0.176)	(0.562)	(0.548)	(0.832)	(1.047)
Dif. STR	-0.098***	-0.050	-0.042	-0.075	-0.151	-0.046
	(0.034)	(0.046)	(0.093)	(0.126)	(0.107)	(0.158)
Dif. Prop. Novice Teachers	-1.555**	-1.379**	-1.229	-0.462	-1.167	0.597
	(0.691)	(0.660)	(1.910)	(2.351)	(2.756)	(3.131)
Cost School of Origin	-0.005***	-0.007***	-0.002	-0.003	-0.005	-0.008
	(0.001)	(0.003)	(0.003)	(0.009)	(0.007)	(0.012)
Constant	83.439***	82.276***	78.330***	81.749***	74.388***	80.911***
	(0.793)	(1.080)	(2.537)	(3.297)	(3.859)	(4.714)
Observations	51,484	33,419	9,476	5,986	5,424	3,321
R-squared	0.116	0.175	0.263	0.337	0.317	0.386
FE: Assigned School	X	X	X	X	X	X
Inc. Female	X	X	X	X	X	X
FE: Dist-Gr-Lev	X	X				
FE: Same Top 2			X	X		
FE: Same Top 3					X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 16: The Effect of School Characteristics on the Likelihood of Accepting and Staying at the Assigned School - Robustness Check - Standardized

VARIABLES	(1) Accept	(2) Stay	(3) Accept	(4) Stay	(5) Accept	(6) Stay
Distance (km)	-2.453***	-3.764***	-3.463***	-6.789***	-5.921***	-7.595***
	(0.277)	(0.456)	(1.186)	(1.135)	(1.529)	(1.760)
Dif. Mean Math sc (s)	3.002***	2.413***	4.298***	3.633**	5.857***	4.306*
	(0.354)	(0.446)	(1.005)	(1.421)	(1.473)	(2.382)
Dif. Mean Parent Edu.	1.925***	4.039***	2.520***	6.431***	0.692	7.519***
	(0.390)	(0.447)	(0.890)	(1.635)	(1.628)	(2.031)
Log(Ratio Num St.)	2.618***	4.722***	2.001	7.041***	3.239	9.254***
	(0.652)	(0.825)	(2.376)	(1.895)	(3.495)	(2.578)
Log(Ratio Num St.)^2	-0.117	-1.681***	0.480	-3.565**	0.773	-5.295*
	(0.480)	(0.523)	(1.666)	(1.622)	(2.465)	(2.960)
Dif. STR	-0.920***	-0.470	-0.398	-0.707	-1.421	-0.754
	(0.321)	(0.435)	(0.878)	(1.194)	(1.010)	(1.524)
Dif. Prop. Novice Teachers	-0.587**	-0.520**	-0.463	-0.174	-0.440	-0.394
	(0.261)	(0.249)	(0.720)	(0.887)	(1.040)	(1.146)
Cost School of Origin	-0.991***	-1.443***	-0.344	-0.541	-0.966	-0.608
	(0.262)	(0.509)	(0.665)	(1.734)	(1.490)	(2.606)
Constant	80.045***	75.131***	73.951***	69.688***	70.367***	65.504***
	(0.083)	(0.164)	(0.391)	(0.574)	(0.686)	(0.965)
Observations	51,484	33,419	9,476	5,986	5,424	3,569
R-squared	0.116	0.175	0.263	0.337	0.317	0.370
FE: Assigned School	X	X	X	X	X	X
Inc. Female	X	X	X	X	X	X
FE: Dist-Gr-Lev	X	X				
FE: Same Top 2			X	X		
FE: Same Top 3					X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

A.2.6 Alternative Specification

This section explores an alternative specification similar to the main one. We will also use two dependent variables: first, Accept_{idsgl} is an indicator variable for student i, district d, school s, level l, and grade g, which takes a value of 100 if the student decides to accept her assignment in the first stage and a value of 0 if the student rejects her assignment; and, second, Staying_{idsgl}, which takes a value of 100 if the student decides to stay at her assigned school and a value of 0 if the student moves to another school conditional on having accepted the assignment.

The independent variable of interest, $\mathcal{X}_{sd}$, aims to capture, for school s in district d, the characteristics of potential interest to parents when they make a decision. First, we calculate the distance Dist_{isd} between the school s and student i's home in district d. In addition, we find information on school class sizes as measured by the student-to-teacher ratio, average parental education, average measures of academic performance (standardized math tests), and school size. We also include characteristics that give us an idea of the relevance of teaching quality, such as the proportion of novice teachers (with less than five years of experience). We include the school of origin's cost to capture an approximate measure of income level and a dummy to control for the student's gender, denoted by δ_i. We include district d by grade g and level l fixed effects, denoted by α_{dgl} to control for the systematic differences in the feasible sets of schools offered. Finally, e_{isdgl} represents the error term.

In summary, we will use the same variables as in specification 6, but in levels instead of the differences between the assigned and original schools. The results are in Table 17 and 18.

A notable finding when comparing the results with Table 7 is that the mean math scores in the assigned schools have no significant effect in terms of the overall level, but they do have a significant effect in terms of the difference to the original school. This seems support the idea that parents base their decisions on the reference of the original school.

The impact of the overall and relative parental education levels are broadly similar. However, we see a decrease in the statistical significance as the sets of fixed-effects vary. Notably, in the second stage, the size of the schools has a substantial effect, which is absent in the first stage. Conversely, the relative size of the school has a significant positive effect in both stages.

$$y_{isdgl} = \alpha_{dgl} + \delta_i + \eta \cdot \text{Dist}_{isd} + \mathcal{X}'_{sd}\Omega + e_{isdgl} \tag{10}$$

Table 17: Alternative Results - Levels

VARIABLES	(1) Accept	(2) Stay	(3) Accept	(4) Stay	(5) Accept	(6) Stay
Distance (km)	-1.511***	-2.049***	-2.227***	-3.100***	-2.808***	-3.171***
	(0.138)	(0.220)	(0.453)	(0.462)	(0.560)	(0.640)
Mean Math sc (s)	0.668	2.070	-1.600	7.090**	0.654	8.206**
	(0.771)	(1.535)	(2.541)	(2.917)	(3.625)	(4.071)
Mean Parent Edu.	1.408***	2.399***	2.192*	6.327***	1.908	7.244***
	(0.490)	(0.827)	(1.259)	(1.672)	(1.911)	(2.669)
Log(Num Students)	5.353	19.473***	-7.548	25.445**	15.115	54.064***
	(4.798)	(7.132)	(16.571)	(12.585)	(20.015)	(19.025)
Log(Num Students)^2	-0.382	-1.290**	0.669	-1.962**	-1.308	-4.188***
	(0.386)	(0.560)	(1.318)	(0.957)	(1.576)	(1.419)
STR	-0.055	0.190**	0.027	0.227	0.058	0.028
	(0.050)	(0.090)	(0.173)	(0.209)	(0.247)	(0.295)
Prop. Novice Teachers	1.828	-1.590	8.028***	7.171*	12.118***	7.740
	(1.230)	(1.593)	(2.131)	(3.798)	(2.875)	(4.917)
Cost School of Origin	-0.017***	-0.025***	-0.015***	-0.029***	-0.023***	-0.042***
	(0.002)	(0.004)	(0.004)	(0.008)	(0.006)	(0.014)
Constant	63.459***	-0.962	91.349*	-34.905	27.876	-125.365**
	(14.676)	(23.288)	(51.020)	(40.152)	(62.816)	(62.626)
Observations	59,525	39,432	11,448	7,602	6,818	4,480
R-squared	0.051	0.062	0.118	0.136	0.130	0.149
Inc. Female	X	X	X	X	X	X
FE: Dist-Gr-Lev	X	X				
FE: Same Top 2			X	X		
FE: Same Top 3					X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 18: Alternative Results - Levels (Standardized)

VARIABLES	(1) Accept	(2) Stay	(3) Accept	(4) Stay	(5) Accept	(6) Stay
Distance (km)	-3.022***	-4.096***	-4.454***	-6.199***	-5.614***	-6.340***
	(0.277)	(0.439)	(0.907)	(0.923)	(1.119)	(1.280)
Mean Math sc (s)	0.223	0.691	-0.534	2.366**	0.218	2.738**
	(0.257)	(0.512)	(0.848)	(0.973)	(1.210)	(1.358)
Mean Parent Edu.	0.936***	1.594***	1.457*	4.205***	1.268	4.814***
	(0.326)	(0.550)	(0.837)	(1.111)	(1.270)	(1.774)
Log(Num Students)	4.517	16.433***	-6.369	21.472**	12.755	45.624***
	(4.049)	(6.019)	(13.984)	(10.620)	(16.891)	(16.055)
Log(Num Students)^2	-3.665	-12.382**	6.418	-18.831**	-12.560	-40.208***
	(3.707)	(5.380)	(12.651)	(9.191)	(15.129)	(13.619)
STR	-0.302	1.043**	0.148	1.246	0.319	0.155
	(0.277)	(0.493)	(0.951)	(1.150)	(1.359)	(1.621)
Prop. Novice Teachers	0.320	-0.279	1.407***	1.257*	2.124***	1.357
	(0.216)	(0.279)	(0.374)	(0.666)	(0.504)	(0.862)
Cost School of Origin	-3.364***	-5.045***	-3.080***	-5.750***	-4.648***	-8.456***
	(0.492)	(0.833)	(0.795)	(1.688)	(1.176)	(2.698)
Constant	80.910***	74.500***	75.232***	70.102***	72.332***	66.882***
	(0.272)	(0.393)	(0.880)	(0.868)	(1.124)	(1.544)
Observations	59,525	39,432	11,448	7,602	6,818	4,480
R-squared	0.051	0.062	0.118	0.136	0.130	0.149
Inc. Female	X	X	X	X	X	X
FE: Dist-Gr-Lev	X	X				
FE: Same Top 2			X	X		
FE: Same Top 3					X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1