Economics of Education Review 108 (2025) 102685

Contents lists available at ScienceDirect

Economics of Education Review

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

# The pandemic exodus: What drives private-to-public school migration in Peru[☆]

Gregory Elacqua[a,*], Nicolás Figueroa[b], Andrés Fontaine[b], Juan Margitic[a], Carolina Méndez[b]

[a] *Inter-American Development Bank, United States*
[b] *Instituto de Economía, Pontificia Universidad Católica de Chile*

ARTICLE INFO

*JEL classification:*
A20
D12
D83
I28

*Keywords:*
School choice
Centralized assignment systems
Private–public schooling
COVID-19

ABSTRACT

In 2020, the Peruvian Ministry of Education responded to the COVID-19 pandemic and an unprecedented rise in demand for public schools by implementing 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 study, we explore empirically the determinants of both accepting a place and remaining in the assigned public school. Exploiting the randomness in assignment created by oversubscribed schools within the deferred acceptance algorithm, we causally estimate the role of distance in the decision to accept the assignment and explore its impact on the decision to remain. We also shed light on the determinants influencing parental preferences. Our results show that families care about distance to the assigned public school, as well as academic and peer quality with respect to their school of origin. Parents weigh distance to school, academic performance, and peer demographics differently based on their familiarity with these characteristics. Consequently, experiencing a given school environment can alter the significance of specific attributes when later deciding whether to stay in the assigned school. While school quality (measured by math scores) is more influential at the acceptance stage, distance and peer demographics become more important in the decision to remain. Our findings offer valuable insights into how governments can strengthen the supply of public schooling.

## 1. Introduction

In 2020, Peru's education sector faced an unprecedented surge in demand for public school vacancies on the part of students looking to migrate from the private to the public sector. The exodus of more than one hundred thousand private-school students, equivalent to approximately 5% of total private enrollment, was likely fueled by a combination of factors: extended school closures due to COVID-19, lack of adaptability of private schools to online education,[1] and the unwavering cost of paying for private education during an economic crisis. Notably, similar trends were observed in other countries in the region (Elacqua, Méndez, & Navarro, 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 2020]) program. This new process used a deferred acceptance (DA) algorithm with the unique feature that parents were unable to declare their preferences. Instead, pseudo-preferences were determined based solely on distance, with the closest school being considered the first preference, the second closest the second preference, and so forth. In this way, the DA algorithm assigned students to schools, randomizing the assignment when schools were oversubscribed. Once the assignments were published, parents could choose to accept or reject the assigned school.

☆ We thank the Ministry of Education of Peru for providing the data for this study. Diana Vásquez, Benjamín Madariaga, and Jadira Sánchez contributed excellent research assistance. We gratefully acknowledge the Inter-American Development Bank (IDB), United States for funding the research. Nicolás Figueroa acknowledges support of ANID PIA/PUENTE AFB230002. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the IDB, 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.

* Corresponding author.
*E-mail address:* gregorye@iadb.org (G. Elacqua).

1 In the Peruvian context, a significant share of private schools are low-fee institutions with limited financial and technological capacity. During the pandemic, many of these schools faced major barriers to transitioning to online education, including lack of infrastructure, limited digital teacher training, and financial stress. By contrast, public schools benefited from centralized support through the national Aprendo en Casa program, which offered multi-platform educational content (radio, TV, and online).

https://doi.org/10.1016/j.econedurev.2025.102685
Received 17 June 2024; Received in revised form 7 July 2025; Accepted 9 July 2025
Available online 19 August 2025

We leverage this random assignment to gauge the impact of distance to the assigned school on two aspects: the probability of families accepting the assignment and, once accepted, the likelihood of the student remaining in the school a year later. The literature on parental choice in the context of education underscores the importance of distance in decision-making (Harris & Larsen, 2017; Hastings, Kane, & Staiger, 2005; Hofflinger, Gelber, & Cañas, 2020; Jacob & Lefgren, 2007). Our analysis extends beyond this work, offering insights into various determinants that influence family decisions to either accept or decline the assigned public schools, and their subsequent choice to remain or switch schools, provided the initial assignment was accepted. We assess various observable attributes, including school size, class size (measured by the student-to-teacher ratio), math test scores, peer demographics, and teacher experience.[2] For those families that transitioned to the public sector, our study evaluates the valuation of specific public school attributes after a year of direct experience, compared to their initial perceptions. While we abstain from making causal inferences based on these attributes, our findings do reveal the factors that families consider relevant. Additionally, we investigate changes in the intensity of these preferences after having experienced the new school environment.

Our research uncovers intricate 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. The unique setting under analysis allows us to identify dimensions of public education valued by parents who previously chose private education, thus helping governments to dispel misconceptions about the public education system.[3] Strengthening public education systems has many potential benefits: they function as a safety net for all members of society (Singh, Park, & Dercon, 2014), as a large-scale adaptable apparatus that allows all students to cope in times of need (such as pandemics, or climate-change induced schooling disruptions), and as an equalizer of inequitable family backgrounds by reducing economic segregation (Balarin, 2015; Elacqua, 2012; Murillo, Belavi, Pinilla et al., 2018).

Our paper first introduces a theoretical framework, which serves as a foundation for our empirical analysis of parental public education preferences. Specifically, we employ a two-period model. In the first period, students are presented with the opportunity to either accept or reject their initial public school assignment, while in the second period, they decide whether or not to remain at their assigned school. In this setting, families observe an additional value of attending a public school and "experimenting", since there is an option value. If they discover that the school is a good fit, they can continue attending, 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 that to continue attending it may appear differently to an econometrician. When deciding whether or not to accept the offer, families are thinking long-term; school characteristics thus play a crucial role since they can have a lasting impact. However, when deciding whether or not to stay in the public school, there is selection bias, since parents who accepted the initial offer may be more sensitive to school characteristics. The model effectively addresses key empirical findings, especially observed shifts in preferences related to essential attributes, such as mathematics test scores and geographical distance.

Our results reveal that proximity to home significantly influences the likelihood of families accepting the assigned public school. It also affects their decision to remain. Specifically, for every additional kilometer increase in the distance to the school, the probability of accepting the assigned school decreases by 2.03 percentage points. The probability of remaining at the same school decreases even more: 3.65 percentage points for every additional kilometer. This suggests that families prioritize convenience and accessibility in their decision-making process, and that this factor is even more important in deciding whether to stay after having experienced the school environment.

We also find suggestive evidence that families are more inclined to accept school assignments when parents' average level of education at the assigned school represents an improvement over the school of origin. Families thus seem to prefer schools with a higher-educated parent community, possibly considering it an indicator of overall school quality. Indeed, parents' preferences are influenced by school quality, as measured by student achievement. Higher standardized math scores (when compared to the school of origin) positively impact the likelihood of accepting the assigned public school. This suggests that parents use the school of origin as a benchmark when making their decisions. In addition, we observe that parents are more likely to accept schools with more experienced teachers, a larger student body, and lower student–teacher ratios (STR). Though, this last outcome does not hold up when more stringent fixed effects are included in the analysis.

With regard to the likelihood of students remaining after their experience attending the assigned school, we cannot establish causality given that the sample consists of families who previously accepted the assignment. However, as mentioned, our results indicate that distance to the school negatively impacts the probability of continued enrollment. Additionally, factors such as the relative demographics of peers (parents' average level of education), school quality (math scores), and the size of the student body compared to the previous school seem to play a significant role in this decision.

Our analysis highlights that families who chose to stay in the public system after one year differ systematically from those who initially accepted their assignment. This distinction allows us to explore how each group prioritizes different school attributes. While we refrain from interpreting these differences as causal effects of experience, the comparison reveals that the two decisions — accepting and staying — reflect the views of two separate populations. For example, although academic quality and peer composition matter in both stages, the relative importance of these factors shifts: school performance (math scores) plays a stronger role at the acceptance stage, whereas after a year in the system, distance and peer demographics (mean parental education) become more influential in the stay decision. This pattern is in line with our theoretical framework, where the stay-or-leave decision is made by a positively selected group that assigns greater weight to dimensions of quality observed after enrollment. As such, while we do not claim that preferences evolve due to experience, our findings offer policy-relevant insights into what different types of families value when engaging with public education.

Our study intersects with the literature on family preferences for schools, particularly work that examines 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 different school characteristics such as academic quality, proximity, and religious instruction. They find these factors are prioritized differently according to socioeconomic background. Parents of lower socioeconomic status tend to prioritize proximity, while those of higher socioeconomic status prioritize student achievement and religious education. Similar results are reported by Abdulkadiroğlu, Pathak, Schellenberg, and Walters (2020), who observe that parents generally prefer schools with higher test scores.

[2] Information about school quality and composition is publicly available through sources such as the School Census (*Censo Escolar*), which includes data on infrastructure, enrollment, and staffing for both public and private schools. Additionally, MINEDU provides access to standardized test results (ECE) and the online platform *Identicole*, where families can compare schools based on academic performance, parental education, and available services. Note, however, that in contrast to the more consistent information on public schools in these systems, on private schools — particularly regarding quality and composition — is sometimes less comprehensive or outdated.

[3] Este País, 1 May 1st, 2014

In terms of revealed preferences, Harris and Larsen (2017) conducted a study in New Orleans and observe that parents preferred schools with higher test scores and value-added measures. However, they also valued 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, as well as underline the heterogeneity of preferences across parent demographics.

Our study aligns with previous research findings. Consistent with (Jacob & Lefgren, 2007) and Burgess, Greaves, Vignoles, and Wilson (2015), we find that families that have been in the private sector prefer higher-performing public schools with favorable peer demographics and proximity to their residence. Furthermore, our results support the observations of Harris and Larsen (2017) regarding consideration of both academic and non-academic attributes in school selection, as well as those of Glazerman and Dotter (2017) concerning the role of commuting distance, peer demographics, and academic indicators.

Our paper also contributes to research exploiting information on parents' preferences and the randomization induced by centralized assignment mechanisms to learn about school choice, such as the that of Abdulkadiroğlu, Agarwal, and Pathak (2017), Glazerman and Dotter (2017), Kapor, Neilson, and Zimmerman (2016), and Agarwal and Somaini (2018). However, our study focuses on a unique case, where preferences in the implemented randomization algorithm were imputed based on distance, in contrast to alternative approaches where parents submit their choices.

The literature reveals that parental preferences for public schools encompass multiple dimensions, including academic quality, proximity, socioeconomic composition, and non-academic factors. Our study adds to this work by examining preferences through a natural experiment, allowing us to capture the perspectives of families who transitioned from private to public schools. In doing so, we offer a unique perspective that can be valuable for governments striving to improve public education, particularly as this concerns evaluating the preferences of families who switched for pandemic-related reasons. Primarily from middle- and low-income backgrounds, they are more broadly representative of a population that has moved from the private to the public sector in recent decades. Better understanding how these families view the public education system and the educational aspects they value is essential for governments to formulate and implement effective policies to strengthen public schooling.

Note that we provide insights on parent preferences based on observed behavior, rather than self-reported data. In general, the stated preference literature (Kleitz, Weiher, Tedin, & Matland, 2000; Schneider, Teske, Marshall, & Roch, 1998) suggests that parents 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 presents the context and critical events that led to the ME 2020 process, as well as the data employed in our study. Section 3 introduces the theoretical model that guides our empirical analysis. Section 4 describes the identification strategy, main results, and a non-linear analysis. Finally, Section 5 discusses the findings and concludes.

## 2. Institutional background

The pandemic caused by COVID-19 led governments throughout Latin America to close schools due to the increased risk of contagion and lack of available vaccines. Countries in the region closed schools for an average of 158 days, in stark contrast with the OECD average of 57 days.[4] There were critical differences in schools' capacity to transition to this new situation, with many facing constraints such as the availability of technological resources and teachers' ability to adapt to remote instruction.[5] Students, meanwhile, were limited in the resources they had access to at home to engage in online instruction, such as technological devices, internet, a physical space to study, etc. Abiz et al. (2022).

The pandemic had severe consequences in Peru, which had the world's highest COVID-19 death rate (far higher than any of its neighbors and twice that of the United States): nearly 6,470 deaths for every 1 million people.[6] Unemployment increased considerably. During the second quarter of 2020, according to the National Institute of Statistics (*Instituto Nacional de Estadística e Informática* - INEI), informal dependent and independent workers were the most affected, registering declines of 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. Sectors like fishing, tourism, construction, commerce, and services experienced the sharpest contractions. Informal employment and small enterprises, especially those with fewer than ten employees, were particularly impacted (Lavado, 2021). Non-primary sectors were also hit hard: employment declined by 67.9% in construction, 58.2% in manufacturing, 56.6% in services, and 54.5% in commerce during the pandemic.

Detailed data on the employment sectors of families in private schools is not available, making it difficult to correlate the regional concentration of these effects with private school enrollment. However, since private enrollment is higher in urban areas — where the most affected economic sectors are concentrated — these places likely faced greater pressure from families seeking to transfer to the public system due to pandemic-related job losses. Rural regions, by contrast, have lower private school enrollment and rely more heavily on public education.

Private schools were reticent to lower their fees,[7] and even threatened to expel students in extreme cases.[8] This reluctance, likely due to a combination of the aforementioned factors and the economic hardships brought about by the pandemic, resulted in significant numbers of students seeking transfers from private to public schools.[9] To address this situation, the government designed a system that could accommodate numerous transfer requests while preserving the principles of equity and efficiency. Specifically, the Ministry of Education (MINEDU), in collaboration with the Education Division at the Inter-American Development Bank, implemented a centralized assignment mechanism. This mechanism utilized the Deferred Acceptance (DA) algorithm to match students with the available vacancies closest to their home.

### 2.1. The exceptional enrollment program in Peru - ME 2020

In Peru, the school year has traditionally followed a calendar that starts in March and ends in December. This structure applies to both

[4] Source: own elaboration based on UNICEF. Period: 11/3/2020 - 2/2/2021.

[5] Infobae, 28 July 2022

[6] Our World Data

[7] El Comercio, 26 April 2020

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

[9] There is no direct data linking the employment status of parents specifically to private school enrollment. However, many families with children in private schools are middle-income households (see Appendix A.2.1) engaged in sectors like commerce, services, construction, and small businesses — all heavily impacted by the pandemic. Informal workers, self-employed individuals, and those running small enterprises faced particularly high levels of job and income loss. While we cannot quantify the exact likelihood, it is reasonable to assume that a significant share of these families experienced employment disruptions, which likely contributed to their inability to continue paying tuition.

public and private schools across the country. Under normal conditions, the enrollment process begins at least one month prior to the start of the school year, with schools providing information about available spots and requirements approximately two months in advance. Families are generally required to complete the process in person at each school. In urban areas — especially Lima — popular schools often attract long queues, sometimes forming days before the enrollment window officially opens. Moreover, there are no standardized or transparent criteria for selection when demand exceeds supply, making the process highly discretionary and unpredictable.

The COVID-19 pandemic disrupted both the school calendar and the enrollment process during the 2020 and 2021 school years. Classes shifted to remote instruction, and in-person visits to schools were restricted, complicating enrollment for both new students — some of whom typically register in March — and those seeking to transfer between schools.

Before the pandemic, some families in Peru opted for private schools due to concerns over public school quality and a belief that private institutions better prepare students for university entrance exams. As documented by Balarin, Kitmang, Ñopo, and Rodríguez (2018), many private schools emphasized content-heavy instruction, particularly in mathematics and science, and adapted their offerings to parental demand for academic rigor. This trend unfolded in a context of a perceived deterioration of public education following economic and policy crises in the 1980s and 1990s. Private enrollment grew from 14% in 1997 to 28% by 2019 nationally, and stood at 52% in Lima. Despite evidence that low-cost private schools often underperform compared to urban public schools, many families — especially those in informal settlements — continued to see private education as the most preferred route to higher education.

In response to the increased demand for public school placements — driven in part by economic shocks that forced families to leave the private system — the MINEDU launched the Exceptional Enrollment Program (*Matrícula Excepcional*, ME) on May 22, 2020. The program aimed to ensure access to public education through a centralized, transparent assignment process.

To facilitate applications, families could submit relevant student information through a web portal, including grade, age, and an address (either residential or another reference location).[10] The ME utilized a Deferred Acceptance algorithm that automatically generated school preferences based on proximity within a five-kilometer radius, as opposed to allowing families to rank schools directly. The nearest school was assigned as the first preference, followed by the next closest, and so on.[11] A sample of these auto-generated preferences is shown in Table 1. The algorithm also gave priority to students with siblings enrolled in a given public school[12] and to students with special educational needs arising from disabilities. There was public information available about school quality and composition through sources such as the School Census data, which includes both public and private schools. This dataset provided information on infrastructure, teacher profiles, educational coverage, and socioeconomic conditions. Additionally, the Ministry of Education manages a platform called Identicole, where parents can access information on enrollment, academic performance, and other relevant aspects of schools.

**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 |

However, while basic information was available for both public and private schools, more detailed indicators of school quality and peer composition are not always publicly accessible, particularly for private schools.

Families then had from 7 June to 21 June to review their assignment. They were free to accept or decline the offered school. If a family declined due to incorrect address information, they could update their data and reapply. The algorithm ran in three rounds to accommodate such corrections and to maximize placement coverage.

Applications were categorized into two types: "first entry" (for students new to the education system) and "transfer" (for students moving from private or other public schools). The MINEDU also offered three kinds of slots: regular, extended, and virtual. The first two refer to an in-person format, but ultimately all instruction in 2020 and 2021 was virtual due to public health conditions.

A total of 125,295 students applied for a spot in a public school through the ME process.[13] Of these, 106,863 accepted the offered place, while 2889 students had various issues with the assignment process and remained pending placement.

## 2.2. Data

Our starting sample includes the 125,295 students who participated in the ME 2020. The students were mainly enrolled in private schools in urban settings[14] and likely submitted an application in response to the shock imposed by the COVID-19 pandemic, which saw extended school closures, an inability of private schools to adapt to online education, and continuing private school tuition fees.

We use data provided by the MINEDU on student background, parent choices, and school characteristics to investigate the factors that determine family decisions. Information on students' enrollment status and teacher characteristics from 2019 to 2021 were obtained 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). The data on average school characteristics comes from multiple sources: the Student Census Evaluation

[10] The platform was disseminated through communication channels at national, regional, and local levels of government. In the months leading up to its launch, private–public sector changes were minimal. The sudden onset of the pandemic had led to widespread closures and significant uncertainty about how long these would last, limiting families' ability to change sectors. If any transfers occurred, they were likely restricted to specific cases handled by phone or other exceptional means. Therefore, it is unlikely that a significant number of students had already changed sectors before the mechanism was introduced, thereby minimizing potential distortions to the sibling priority criterion.

[11] Schools needed to have available spots in order to be included as an option in the allocation mechanism. Therefore, a critical input for the platform was each school's declaration of available vacancies. Schools were required to report the number of open seats, specifying the corresponding grade levels and time shifts (morning or afternoon).

[12] Approximately 29% of the applicants in the ME 2020 process had at least one sibling already enrolled in the assigned school, and thus benefited from this priority rule.

[13] It represents approximately 5% of total private enrollment in Peru prior to the pandemic.

[14] The vast majority of private schools — and therefore of students who applied through the ME 2020 process — are located in urban areas. Rural private schools are much less common in Peru, and consequently, the number of students in our sample coming from rural areas is quite small. Specifically, approximately 1.5% of the private schools attended by students who participated in the process are located in rural areas. Within urban areas, private schools are more concentrated in densely populated districts, especially in Lima and other major cities.

(*Evaluacion Censal de Estudiantes*, ECE) for measures of parent education and schools' standardized academic performance from 2014–2019, and the School Directory (*Padron Escolar*) and School Census (*Censo Escolar*) for information on staff and facilities from 2014–2021.

### 2.3. Sample definition

We are interested in analyzing two distinct stages of the process. The first involves parents either accepting or declining the assigned public school for their children. The second stage consists of families deciding whether to remain or leave the assigned and accepted public school. For both stages, we exclude students who participated in the second and third rounds of school assignments. These students faced different pools of potential schools for assignment, and some of their pseudo-preferences were inaccurately presumed due to incorrect address information. Additionally, we focus our analysis exclusively on students seeking to leave private schools, as our aim is to examine a population that had previously opted for the private sector over the public sector.[15]

We further restrict our sample to students who in 2021 were attending a school within a 50-kilometer radius of the school to which they were assigned through the ME 2020 process. We recognize that students may switch schools for reasons unrelated to the schools' characteristics, such as family relocation. Our analysis accordingly includes only those students who were enrolled in a school near their original assignment in 2021[16].

To ensure an authentic random distribution in our process, we excluded clusters where each student had been assigned their preferred school, indicating that randomization had not been necessary, as everyone had received their first choice. In contrast, students ending up with their second choice or lower is evidence that randomization had indeed been employed. We identified these clusters through two methods. We grouped students into a single cluster if they were from the same district, same education level, and same grade. For a more rigorous identification, we formed clusters of students who had the same top two school ranked on their list, based on proximity to their home address.

In the second stage, we exclude students who declined their assignment,[17] as they did not experience the environment of their assigned public school. Recall that our goal is to identify the factors that determine a student's decision to stay in a public school after having experienced it firsthand.

We define *accepting* the school assignment if the student accepted and virtually attended their assigned school. Note that some students, despite having accepted their designated school, did not proceed with enrollment.[18] For these cases, we assume that these students *de facto* rejected their assignment, and thus re-code their responses in the first stage. We also observed that some students accepted their school assignment, but were last recorded in 2020 as being registered at a different school after the ME allocation. If, for example a student accepts a placement from the ME 2020 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, having potentially started there in June, before deciding to switch by September.[19] Thus, "treated" students are those who were exposed to

**Table 2**
Number of observations lost at each stage of the sample selection.

| Stage sample filtration | Num. observations lost. |
|---|---|
| School in 2021 more than 50 km away | 23,670 |
| Exclude round 2 and 3 | 12,564 |
| Students come from public school | 12,097 |

the assigned public school. After applying our sampling criteria, we retained 61.43% (or 76,964 students, with 13,556 of them grouped by their top two ranked schools)) of the original pool of students in the first stage of the ME 2020 process. In the second stage, our sample comprises 46.59% (or 58,377 students, with 12,903 grouped by their top two ranked schools) of the original population.[20] Fig. 1 illustrates the progression of students (all from the private sector) from the first to the second stage. The horizontal dotted line divides the sample into two stages. The green arrows represent students' decisions to pursue their education in the school allocated during the ME 2020 process. The red arrows symbolize the choice to decline the placement. The first row displays initial decisions to either accept or decline the ME 2020 school assignment. The second row indicates those students who accepted their placement but never actually enrolled in the assigned school. We label these as *"de-facto"* rejections. The third and fourth rows map out student decisions to transfer after spending a few months at their assigned school in 2020 and the subsequent decision made by March 2021.

### 2.4. Sample description

In this section, we offer a detailed examination of the traits of schools participating in the ME 2020 process.

We compare private schools from the entire school registry in Peru (referred to as the "Population of schools in Peru") to the schools of origin of students who participated in the ME 2020 process (which we term the "Full Sample"). This is further compared to schools included in our final dataset, the "Regression Sample". Table 3 illustrates the distribution across school grades in the different samples. Fig. 2 and Panel A of Fig. 3 also display the distributions of select characteristics. These illustrations demonstrate that, even after applying our selection criteria, our sample remains representative of the broader school population.

We also conduct a comparative analysis of public schools. Here, we contrast the attributes of all public schools in the registry to those schools assigned to students during the ME 2020 process ("Full Sample") and those present in our "Regression Sample". As illustrated in Fig. 4 and Panel B of Fig. 3, subtle variations emerge in the distributions. The assigned schools exhibit higher levels of parental education, larger school sizes, and a lower prevalence of institutions with a very low socioeconomic index compared to public schools in the broader population. This observation suggests that applicants were not generally placed in the most disadvantaged schools.

Narrowing our analysis to the regression sample, we examine the evolution of covariates between stages, presenting a detailed description in Table 4. Surprisingly, we observe no substantial changes between the stages, contrary to initial expectations. This suggests a notable stability in the considered factors throughout the analyzed stages, challenging preconceived notions. Panel C of Table 4 displays differences in the covariates between the assigned school and the school of origin. In this context, we observe more significant changes. Table 5 shows the same results but considering all the filtered sample.

[15] 8487 of our sample were enrolled in public schools.

[16] See Table 2 for detail in the number of observations lost at each stage of sample selection

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

[18] In Fig. 1, they represent 5579 (5485 + 94) of the 69,541 students who initially accepted their assignment.

[19] In Fig. 1, they represent the 2091 students who migrated to private schools in 2020 and the 2801 students who migrated to a different public school among the 57,496 students who were effectively "treated".

[20] It is important to note a disparity between our regression samples and the figures we report, which arises primarily for two reasons. First, our preferred approach involved breaking down the samples based on top two ranked school. However, this meant that many subsets lacked sufficient variation in decisions to accept or decline school assignments. Second, missing data for specific students on certain analytical factors posed a challenge, though this only impacted a relatively small portion.

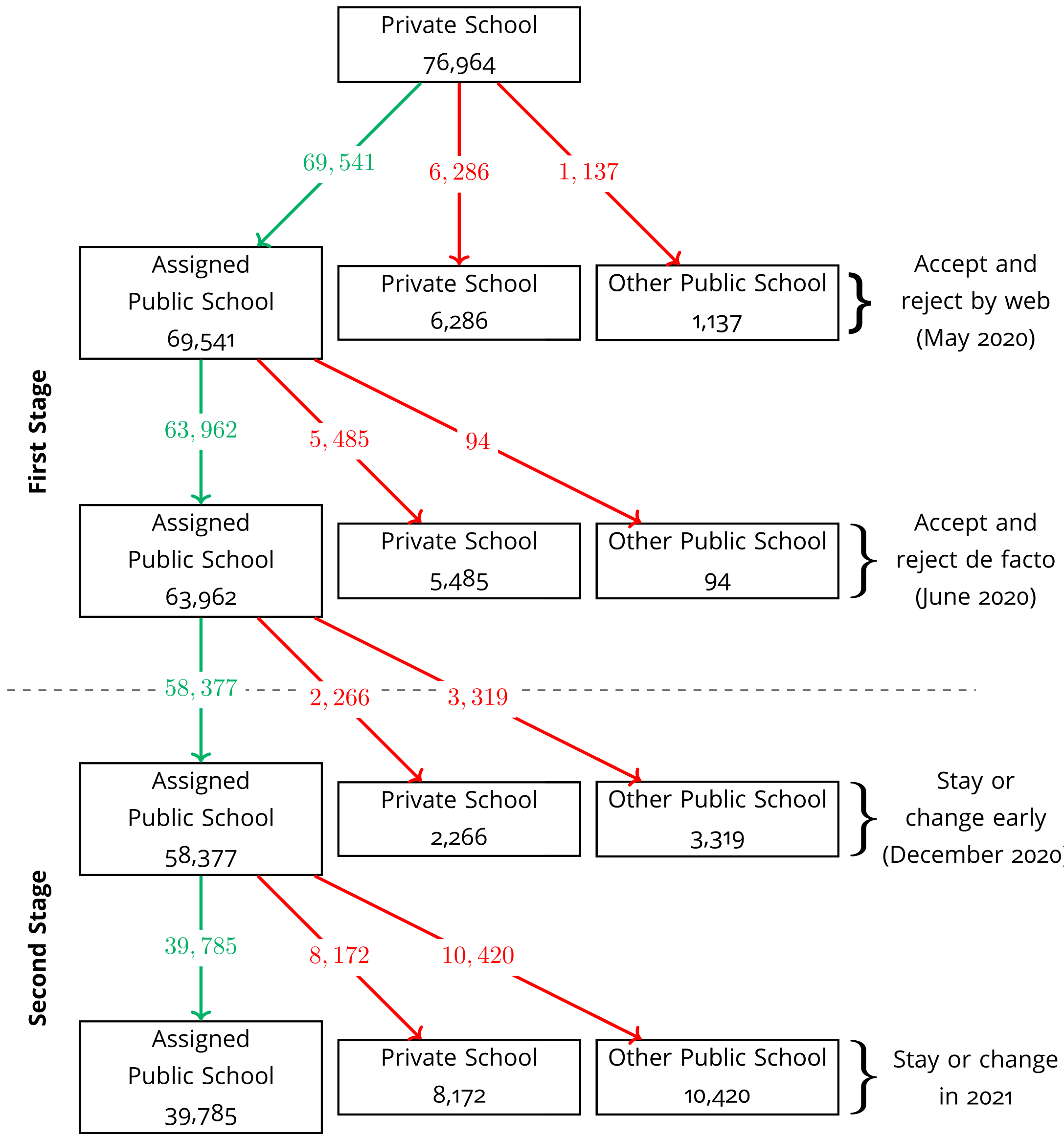



**Fig. 1.** Student flow before and after the ME 2020.

For instance, students who accepted their assignment exhibit a larger disparity in academic performance in mathematics. That is, those who accepted the public school placement went to schools with higher average scores in mathematics compared to the schools of origin. Furthermore, students who decided to remain in their assigned school are characterized by even higher average scores. A similar pattern emerges with parental education. While the assigned schools had, on average, lower parental educational levels compared to the schools of origin, the gap narrows when considering students who both accepted and remained in the assigned school.[21] Additionally, the assigned schools where students chose to accept a place and remain tend to have larger student populations than the schools of origin.

Interestingly, our analysis reveals that while there are minimal alterations in the covariates at education levels between the different stages, pronounced disparities emerge in terms of differences between assigned schools and schools of origin. This would suggest that parents make decisions based on their prior experience with a school, and that with the new school assignment, their experience may change significantly.

## 3. Theoretical framework

We now introduce a two-period model that allows us to rationalize the empirical findings in Section 4.2. In the first period, students decide

[21] Parental education is measured as a categorical variable and represents different levels of educational attainment by parents. These are averaged at the school level based on information from the ECE dataset. The numerical values are as follows:

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

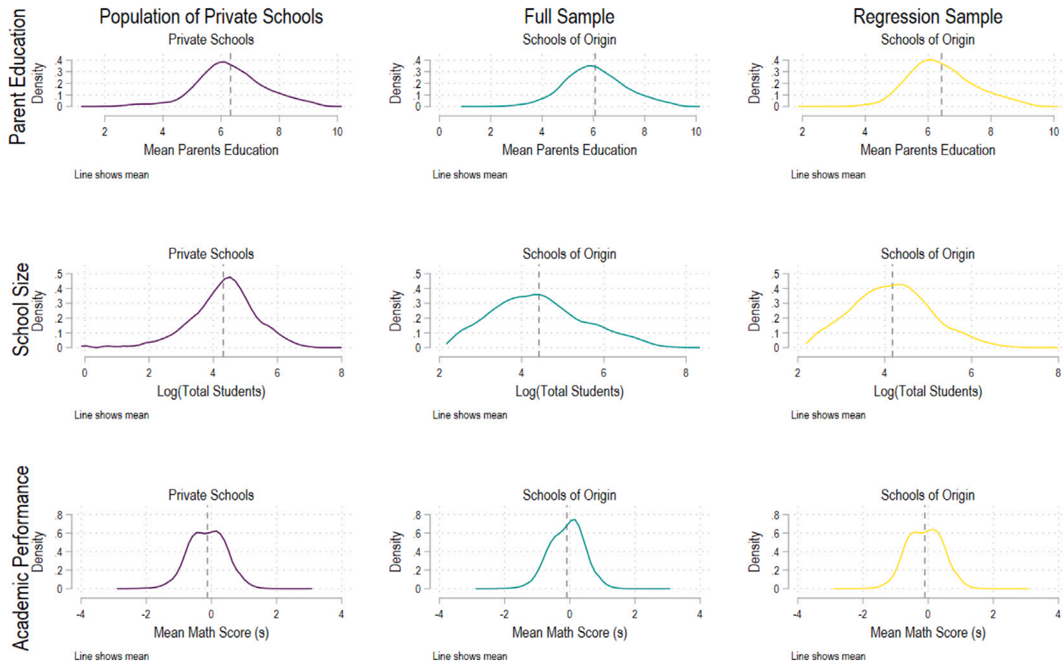



**Fig. 2.** Comparative distribution of private school characteristics among population, full sample, and regression sample.
Note: Each column represents a different sample. The first includes all private schools in Peru; the second includes only the private schools attended by students who participated in the ME 2020 process; and the third corresponds to the private schools of students remaining after all filtering steps. Parental education is measured using categorical levels, school size is measured as the logarithm of total enrollment, and academic performance is measured as the standardized average score on a mathematics assessment.

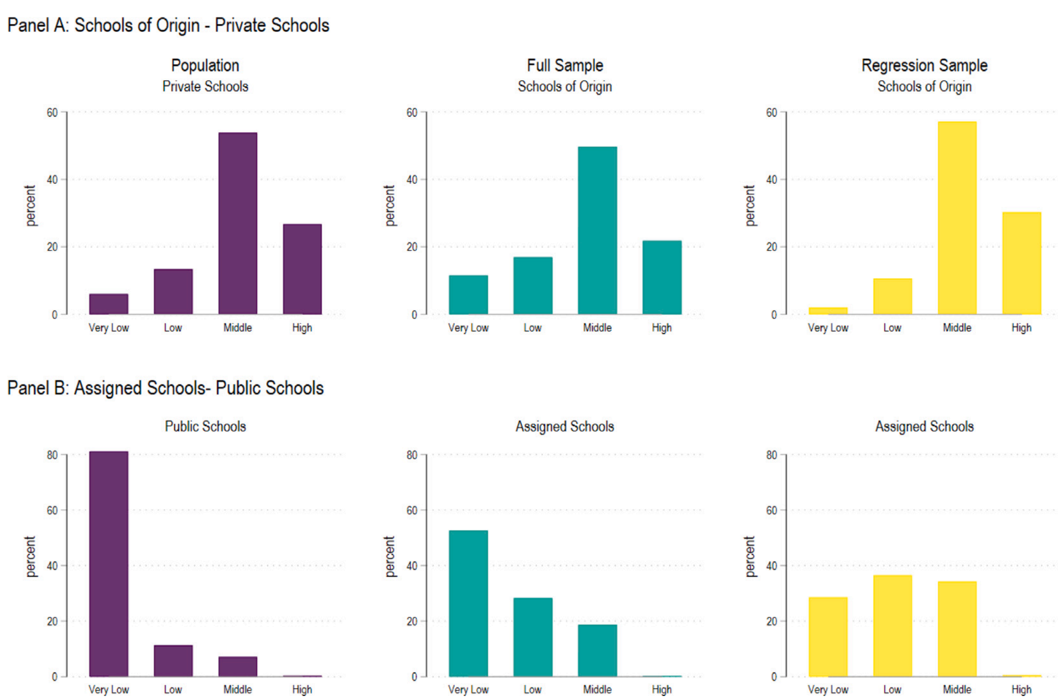



**Fig. 3.** Distributions of socioeconomic index in private and public schools.
Note: Panel A corresponds to private schools, and Panel B to public schools. Each column represents a different sample. The first includes all private/public schools in Peru; the second includes only the private/public schools attended by students who participated in the ME 2020 process; and the third includes the private/public schools of students remaining after all filtering steps.

whether or not to switch to the public sector based on the school they are offered. In this initial decision, students are imperfectly informed about the offered school. 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 component and decide whether to remain or leave. Students fully internalize the "experimentation value" of switching to a public school: they can stay if it is a good fit or return to a private school if it is 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 or not to enroll in a public school, the school's characteristics play a pivotal role since


**Fig. 4.** Comparative distribution of public school characteristics among population, full sample, and regression sample.
Note: Each column represents a different sample. The first includes all public schools in Peru; the second includes only the public schools attended by students who participated in the ME 2020 process; and the third corresponds to the public schools of students remaining after all filtering steps. Parental education is measured using categorical levels, school size is measured as the logarithm of total enrollment, and academic performance is measured as the standardized average score on a mathematics assessment.

**Table 3**
Distribution across school grades in 2020.

| Grade | Full sample | Regression sample | Acceptance sample |
|---|---|---|---|
| **Initial** | 27,946 | 16,243 | 13,962 |
| **Primary** | | | |
| 1 | 13,166 | 8457 | 6969 |
| 2 | 12,064 | 8853 | 7162 |
| 3 | 10,427 | 7421 | 6019 |
| 4 | 9261 | 6512 | 5340 |
| 5 | 8700 | 5921 | 4757 |
| 6 | 7687 | 5417 | 4263 |
| Total | 61,305 | 42,581 | 34,510 |
| **Secondary** | | | |
| 1 | 7497 | 4328 | 3505 |
| 2 | 6403 | 4078 | 3287 |
| 3 | 6168 | 3808 | 3080 |
| 4 | 6049 | 3926 | 3136 |
| 5 | 5552 | 3588 | 2917 |
| Total | 31,669 | 19,728 | 15,925 |

(a) Note: The full sample includes all students who participated in the ME 2020. The regression sample consists of students remaining after the filtering criteria specified above. Finally, the acceptance sample includes only those students who accepted their assignments.

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 it comes to deciding whether or not to continue at the school, the choice is made by a specific subset: those who initially decided to try the public school. This group may place greater emphasis on these school characteristics. The magnitude of these competing effects will determine whether decisions in the first or second stage are more influenced by school attributes.

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 period 1 due to COVID, which reduces their valuation for the private school, either because of economic difficulties or a decrease in service quality. Additionally, $\varepsilon$ and $\eta$ 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 ($\varepsilon$) and the strength of recovery after the crisis ($\eta$). Agents know the realization of $\varepsilon$ before making their decision at $t = 1$ and the realization of $\eta$ 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 $\alpha V_{PU}$, where $V_{PU}$ is an objective measure of school attractiveness capturing observable attributes such as distance, parents' education, or schools' results on standardized tests. In its empirical counterpart, $V_{PU}$ is represented by the vector of observable attributes of a given school.

On the other hand, $\alpha$ 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 $\alpha$, 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 $\mu_0$ and $1-\mu_0$. The distribution of $\alpha$ is specified as $F_i$ on $[1, K]$, with $K < \infty$. We assume that $F_1 \succeq_{FOSD} F_2$, signifying that a fraction $\mu_0$ 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 next assumption captures the idea that the worst possible match with public schools ($\alpha = 1$) makes it undesirable even in period 1, but the best possible match ($\alpha = K$) makes it the best option even in period 2, after the crisis has passed.

**Table 4**
Covariates by stages - Regression Sample.

| | Sample | | | Accept | | | Stay | | |
|---|---|---|---|---|---|---|---|---|---|
| | N | Mean | SD | N | Mean | SD | N | Mean | SD |
| **Panel A: Assigned schools** | | | | | | | | | |
| Distance (km) | 13 552 | 2.09 | 1.41 | 9949 | 2.05 | 1.42 | 6402 | 2.01 | 1.41 |
| STR | 13 552 | 23.08 | 4.77 | 9949 | 23.18 | 4.83 | 6402 | 23.56 | 4.62 |
| Mean Math sc (s) | 11 619 | 0.02 | 0.30 | 8858 | 0.02 | 0.30 | 5777 | 0.02 | 0.30 |
| Mean Parent Edu. | 11 619 | 4.93 | 0.60 | 8857 | 4.93 | 0.57 | 5777 | 4.93 | 0.56 |
| Log (Num. of Students) | 13 552 | 6.20 | 0.69 | 9949 | 6.24 | 0.67 | 6402 | 6.29 | 0.66 |
| Prop. Novice Teachers | 13 482 | 0.14 | 0.17 | 9908 | 0.14 | 0.17 | 6375 | 0.14 | 0.17 |
| Cost School of Origin | 13 426 | 244.62 | 203.81 | 9834 | 235.48 | 188.75 | 6325 | 223.40 | 188.28 |
| **Panel B: Schools of origin** | | | | | | | | | |
| Distance (km) | 13 405 | 2.20 | 3.78 | 9825 | 2.27 | 3.86 | 6321 | 2.21 | 3.85 |
| STR | 13 356 | 17.94 | 9.44 | 9797 | 17.82 | 9.23 | 6300 | 17.62 | 9.07 |
| Mean Math sc (s) | 10 833 | −0.09 | 0.55 | 8150 | −0.12 | 0.54 | 5278 | −0.19 | 0.52 |
| Mean Parent Edu. | 10 615 | 6.47 | 1.03 | 7964 | 6.40 | 0.99 | 5138 | 6.27 | 0.95 |
| Log (Num. of Students) | 13 128 | 4.79 | 0.96 | 9589 | 4.79 | 0.94 | 6152 | 4.72 | 0.92 |
| Prop. Novice Teachers | 12 954 | 0.33 | 0.34 | 9479 | 0.34 | 0.34 | 6096 | 0.34 | 0.35 |
| Cost School of Origin | 13 426 | 244.62 | 203.81 | 9834 | 235.48 | 188.75 | 6325 | 223.40 | 188.28 |
| **Panel C: Differences in covariates** | | | | | | | | | |
| Dif. Distance (km) | 13 402 | −0.10 | 3.97 | 9823 | −0.21 | 4.05 | 6321 | −0.19 | 4.03 |
| Dif. STR | 13 286 | 5.43 | 9.33 | 9765 | 5.54 | 9.53 | 6286 | 6.08 | 9.27 |
| Dif. Mean Math sc (s) | 10 809 | 0.11 | 0.57 | 8126 | 0.14 | 0.56 | 5268 | 0.22 | 0.54 |
| Dif. Mean Parent Edu. | 10 591 | −1.44 | 1.03 | 7940 | −1.37 | 0.97 | 5128 | −1.24 | 0.91 |
| Log (Ratio Num St.) | 13 353 | 1.31 | 1.04 | 9796 | 1.37 | 1.05 | 6301 | 1.47 | 1.01 |
| Dif. Prop Novice Teachers | 12 884 | −0.20 | 0.38 | 9440 | −0.20 | 0.38 | 6071 | −0.20 | 0.38 |

(a) Note: The table reports descriptive statistics of the schools. The statistics are based on the sample that includes groups of students who share the same top two ranked schools, where the top school has full capacity.

**Table 5**
Covariates by stages - Filtered Sample.

| | Sample | | | Accept | | | Stay | | |
|---|---|---|---|---|---|---|---|---|---|
| | N | Mean | SD | N | Mean | SD | N | Mean | SD |
| **Panel A: Assigned schools** | | | | | | | | | |
| Distance (km) | 76 343 | 2.15 | 1.40 | 50 069 | 2.10 | 1.40 | 35 052 | 2.08 | 1.40 |
| STR | 76 324 | 22.48 | 4.94 | 50 056 | 22.57 | 4.97 | 35 044 | 22.75 | 4.91 |
| Mean Math sc (s) | 60 515 | 0.00 | 0.31 | 42 596 | −0.00 | 0.32 | 30 551 | 0.01 | 0.31 |
| Mean Parent Edu. | 60 508 | 4.93 | 0.62 | 42 591 | 4.93 | 0.62 | 30 551 | 4.95 | 0.60 |
| Log (Num. of Students) | 76 324 | 6.03 | 0.75 | 50 056 | 6.10 | 0.73 | 35 044 | 6.16 | 0.70 |
| Prop. Novice Teachers | 75 913 | 0.14 | 0.17 | 49 797 | 0.14 | 0.17 | 34 852 | 0.13 | 0.16 |
| Cost School of Origin | 75 495 | 236.37 | 195.39 | 49 488 | 229.18 | 186.16 | 34 605 | 220.32 | 179.18 |
| **Panel B: Schools of origin** | | | | | | | | | |
| Distance (km) | 74 393 | 2.23 | 3.88 | 48 524 | 2.27 | 3.98 | 34 070 | 2.23 | 3.98 |
| STR | 74 478 | 17.42 | 8.99 | 48 762 | 17.22 | 8.90 | 34 099 | 17.11 | 8.99 |
| Mean Math sc (s) | 56 005 | −0.07 | 0.54 | 38 855 | −0.11 | 0.53 | 27 796 | −0.14 | 0.53 |
| Mean Parent Edu. | 54 975 | 6.49 | 0.99 | 38 024 | 6.42 | 0.96 | 27 152 | 6.35 | 0.93 |
| Log (Num. of Students) | 72 947 | 4.72 | 0.97 | 47 734 | 4.72 | 0.94 | 33 367 | 4.70 | 0.93 |
| Prop. Novice Teachers | 72 115 | 0.33 | 0.34 | 47 190 | 0.34 | 0.34 | 32 988 | 0.34 | 0.34 |
| Cost School of Origin | 74 808 | 236.78 | 196.04 | 48 975 | 229.59 | 186.91 | 34 234 | 220.72 | 179.98 |
| **Panel C: Differences in covariates** | | | | | | | | | |
| Dif. Distance (km) | 75 047 | −0.07 | 4.06 | 49 019 | −0.16 | 4.16 | 34 441 | −0.15 | 4.15 |
| Dif. STR | 74 795 | 5.29 | 9.28 | 49 055 | 5.54 | 9.33 | 34 338 | 5.82 | 9.35 |
| Dif. Mean Math sc (s) | 56 471 | 0.08 | 0.56 | 39 183 | 0.12 | 0.55 | 28 073 | 0.15 | 0.54 |
| Dif. Mean Parent Edu. | 55 422 | −1.45 | 1.01 | 38 338 | −1.38 | 0.99 | 27 419 | −1.30 | 0.95 |
| Log (Ratio Num St.) | 75 096 | 1.22 | 1.05 | 49 231 | 1.30 | 1.05 | 34 462 | 1.38 | 1.04 |
| Dif. Prop Novice Teachers | 72 362 | −0.20 | 0.38 | 47 427 | −0.20 | 0.38 | 33 174 | −0.20 | 0.38 |

(a) Note: The table reports descriptive statistics of the schools. The statistics are based on the Regression Sample after applying all filtering steps.

**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 having knowledge only of the 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 $\alpha$ 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:

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

or equivalently if

$$\mathbb{E}(\alpha)V_{PU} + \mathbb{E}(max\left\{\alpha V_{PU}, V_{PR}^{(2)}\right\}) \geq V_{PR}^{(1)} + V_{PR}^{(2)} + \varepsilon \tag{2}$$

where $\varepsilon$ 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 $\alpha$ 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:*

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

*In essence, for each type of student, there exists a threshold value for $\varepsilon$, beyond which the student chooses not to try the public school. 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 $F_1 \succeq_{FOSD} F_2$. □ In $t = 2$, students who tried out a public school choose whether or not to stay in the public system. In this period, they already know their match with the public school ($\alpha$) and their valuation of going back to the private sector ($V_{PR}^{(2)} + \eta$). Thus, a student stays in their public school if:

$$\alpha V_{PU} \geq V_{PR}^{(2)} + \eta \tag{3}$$

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

**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:*

$$\mathbb{P}(stay) = \mu_1 \int_{\alpha=1}^{K} G(L(\alpha V_{PU})) d F_1(\alpha) + (1 - \mu_1) \int_{\alpha=1}^{K} G(L(\alpha V_{PU})) d F_2(\alpha)$$

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

**Proof.** Direct. □

By itself, the previous lemma shows two important things with a direct empirical implication. The first, and quite direct, is that as $V_{PU}$ increases, students are more likely to go to the public school they were assigned. Secondly, 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 corresponds to 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 are given by the parameters $\beta_i$ in the regressions, which indicate the increased likelihood of an option given a change in the observable characteristic of a school.

As expected, the likelihood of deciding to try the public school at $t = 1$ and staying in it at $t = 2$ are increasing in $V_{PU}$. What is more interesting is that 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, families already enrolled in the public school system are more sensitive to school quality. On the other hand, 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 then be greater if:*

$$(\mu_1 - \mu_0)(\bar{\alpha}_1 - \bar{\alpha}_2) + \lambda(V_{PU}) > \left[ (1 - \mu_0) \int_{\alpha=\frac{V_{PR}^{(2)}}{V_{PU}}}^{K} \alpha \cdot d F_2(\alpha) + \mu_0 \int_{\alpha=\frac{V_{PR}^{(2)}}{V_{PU}}}^{K} \alpha \cdot d F_1(\alpha) \right] \tag{4}$$

*where $\lambda(V_{PU}) = \frac{\partial \mu_1}{\partial V_{PU}} \left[ \int_{\alpha=1}^{K} G(L(\alpha V_{PU})) d F_1(\alpha) - \int_{\alpha=1}^{K} G(L(\alpha V_{PU})) d F_2(\alpha) \right]$ and $\bar{\alpha}_i = \mathbb{E}_{F_i}(\alpha)$.*

The proof can be found in Appendix A.1

In condition (4), the left-hand side quantifies the magnitude of the selection effect. As "optimistic" students, with higher values of $\alpha$, 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 (4) represents the *horizon effect*. At $t = 1$, $V_{PU}$ assumes greater importance because it has the potential to influence the student for two 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, the horizon effect is purely mechanical in nature, since 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 $H_1$ is more elastic than $H_2$,[22], which is to say $V_{PU}$ induces a bigger percentage change on the school valuation for group 1 than for group 2. Additionally, $\mu_1 > \mu_0$, as in the second stage, only those 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 the assignment 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 decision — whether or not to accept the assigned public school or whether or not 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 important for students' decisions *after* they have already had a year of experience in the public school. The findings in Section 4.1.3 align with a strong selection effect. In addition, they suggest that families 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 analysis into three subsections. First, we introduce the identification strategy for our causal estimation of the effect of distance, along with the framework for our exploratory analysis

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

of the determinants of acceptance and permanence in public education. The second and third subsections present the main results of the role of distance, as well as the other determinants of parental choices, respectively.

### 4.1. Identification strategy

The ME 2020 algorithmic allocation process allows us to exploit exogenous variation from the randomized assignment of students conditional on the local school offering to recover the causal estimates of the parents' preference 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 school proximity to the student's home or address of her preference (ranking the closest available school first and the furthest last) within a 5 km radius.

Conditional on families ranking public schools by proximity, students are randomly allocated to public schools to maximize the likelihood of getting a spot at their most preferred school (in this context, closest to their home). Therefore, students with similar ranking, 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 education level, no randomization will take place.[23]

Our identification strategy relies on the exogeneity on the assignment of schools between students conditional on these students having similar ranked schools (list of closest schools). A possible threat to this assumption would be observing parents strategically listing their addresses close to their preferred school. However, we do not believe this is a concern. See Appendix A.2.2 for a detailed analysis of this concern.

#### 4.1.1. The role of distance

In this subsection, we introduce the main specification used to estimate the causal effect of distance on parents' decisions to accept or reject the assigned school.

$$y_{idsgl} = \alpha_{dgl} + \gamma_s + \delta_i + \beta_1 \cdot \text{Dist}_{is} + \beta_2 \cdot (\text{Price School Origin})_i + e_{idsgl} \quad (5)$$

Where $y_{idsgl}$ takes two values: first, $\text{Accept}_{idsgl}$ is an indicator variable for student $i$ in district $d$, assigned school $s$, grade $g$ and education 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, $\text{Staying}_{idsgl}$, 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 she moves to another school conditional on having accepted the assignment. The independent variables used include $\text{Dist}_{is}$, which is the distance between the assigned school $s$ and the address of student $i$. We include the price 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 $\delta_i$. Given that the ideal experiment would compare two identical students (including those who attended the same school at baseline) with identical assigned-school rankings (e.g., closest, second-closest, third-closest school), but where one is randomly assigned to a more distant school, we incorporate granular fixed effects. Specifically, we restrict comparisons to students with the same top two schools listed in their ranking (based on proximity). Comparing students with the same schools listed among their top priorities on oversubscribed schools closely mimics the randomization mechanism used by the algorithm.[24]

[23] It is for this reason that we drop all clusters where students are all assigned to their first ranked school.

[24] We also refined our fixed effects at a more granular level for comparison among those who have the same top 3 schools. These results can be found in Appendix A.2.7.

**Table 6**
Relation between distance and covariates.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay |
|---|---|---|---|---|
| Mean Math sc (s) | 0.189 | 0.453 | 0.157 | 0.141 |
| | (0.318) | (0.335) | (0.167) | (0.166) |
| Mean Lang sc (s) | −0.426 | −0.855* | −0.182 | −0.169 |
| | (0.405) | (0.461) | (0.214) | (0.216) |
| Mean Parent Edu. | 0.032 | 0.067 | −0.039 | −0.016 |
| | (0.108) | (0.135) | (0.058) | (0.067) |
| Log (Num Students) | −0.499 | −0.982 | −0.537 | −0.469 |
| | (0.787) | (0.838) | (0.388) | (0.399) |
| Log $(\text{Num Students})^2$ | 0.022 | 0.065 | 0.034 | 0.028 |
| | (0.065) | (0.070) | (0.032) | (0.033) |
| STR | −0.016 | −0.017 | 0.002 | 0.002 |
| | (0.013) | (0.015) | (0.006) | (0.007) |
| Prop. Novice Teachers | 0.245 | 0.372** | 0.289** | 0.295** |
| | (0.158) | (0.151) | (0.111) | (0.127) |
| Price School of Origin | −0.000 | 0.000 | −0.000 | −0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Constant | 4.517* | 5.639** | 4.261*** | 3.938*** |
| | (2.336) | (2.534) | (1.169) | (1.207) |
| Observations | 11,448 | 7602 | 58,857 | 38,951 |
| R-squared | 0.395 | 0.378 | 0.349 | 0.335 |
| FE: Dist-Gr-Lev | X | X | X | X |
| FE: Female | X | X | X | X |
| F-test: | 6.84 | 6.84 | 6.84 | 6.84 |
| Prob > F: | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

(a) Note: Coefficients correspond to equation (4). In columns (1) and (2) we use same top 2 schools fixed effects and in columns (3) and (4) district-grade-education level fixed effects. Standard errors are clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In addition, we include an assigned school-level fixed effect, $\gamma_s$, which controls for the time invariant characteristics of the school. Finally, $e_{idsgl}$ represents the error term.

In Table 6 we regress distance to school on observable characteristics of the assigned school. The same fixed effect specifications from Eqs. (5) and (7), are retained, except for the fixed effect 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 uncorrelated with unobservable characteristics of the assigned schools. Notably, only the proportion of novice teachers shows statistical significance among the covariates. Nonetheless, the covariates do not exhibit joint significance in their relationship with distance.

Fig. 5 shows the acceptance rate by a school's proximity ranking. We observe a clear negative relationship: families are more likely to accept their assignment when the school ranks higher in terms of proximity.

#### 4.1.2. Non-linear effects of distance

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

$$y_{idsgl} = \alpha_{dgl} + \gamma_s + \delta_i + \sum_{r=2}^{5} \eta_r \cdot I\{Dist_{is} \in Q_r\} + \beta_2 \cdot (\text{Price School Origin})_i + e_{idsgl} \quad (6)$$

As seen in Eq. (6), we relax the linear constraint on modeling the effect of distance. To do this, we employ the specification (5), with a modification regarding the distance variable. Instead of using the continuous distance variable, we replace it with an indicator that separates distances into five categories, each representing a 1 km range: [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 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.

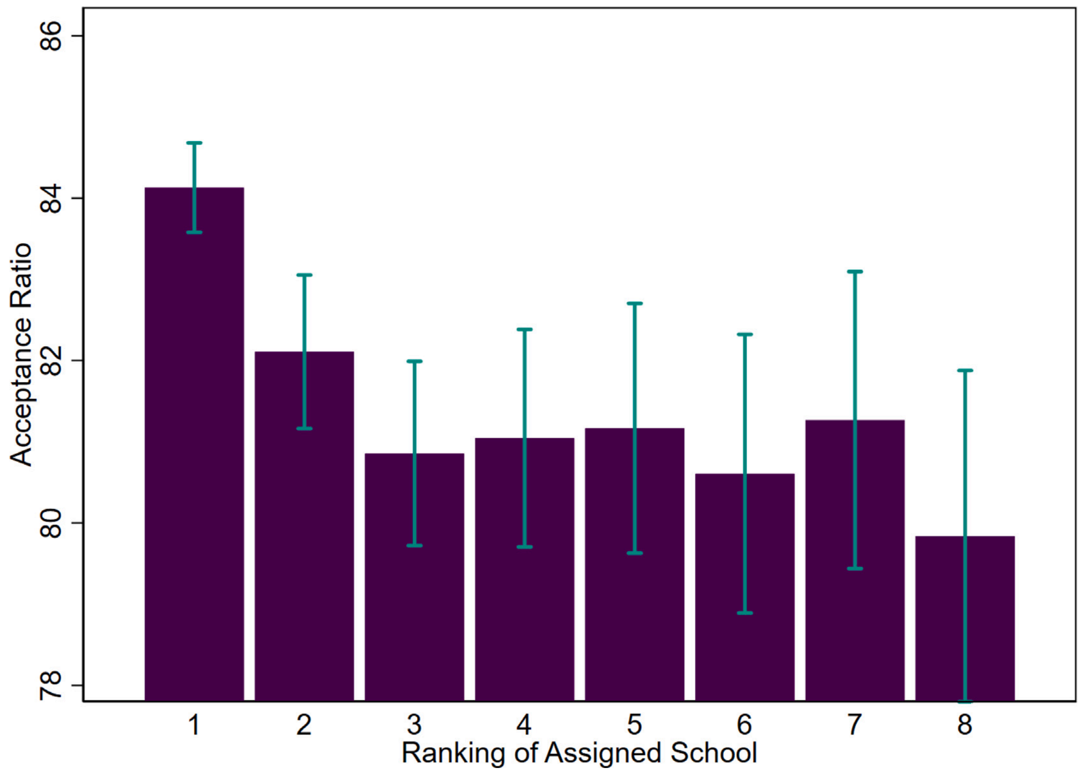



**Fig. 5.** Distribution of acceptances by preference ranking.
Note: The ranking of preferences is constructed such that the first ranked school is the closest school to the student.

Similarly, in Appendix A.2.8, we employ a categorization that divides distance into quartiles based on the distribution of distance to school.

#### 4.1.3. The determinants of parents' preferences

Next, we introduce the specification used to estimate the determinants of parental preferences in their decision to accept the assigned school and stay enrolled after a year. The estimates produced by (7) 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 regarding relative school characteristics in relation to accepting or rejecting their assignment offering. Then, conditional on accepting their assignment in the first stage, we examine the influence of observable characteristics on the decision to remain at or leave the assigned school.[25] In particular, we run the following regression:

$$y_{idsgl} = \alpha_{dgl} + \gamma_s + \delta_i + \psi \cdot \text{Dist}_{is} + \mathcal{D}'_{is}\Lambda + \beta_2 \cdot (\text{Price School Origin})_i + e_{idsgl} \quad (7)$$

Where $y_{isdgl}$ takes two values: first, $\text{Accept}_{idsgl}$ is an indicator variable for student $i$ in district $d$, assigned to school $s$, education level $l$, and grade $g$ that takes a value of 100 if the student decides to accept her assignment in the first stage and a value of 0 if she rejects it; and second, $\text{Staying}_{idsgl}$, which takes a value of 100 if the student decides to stay at her assigned school and a value of 0 if she moves to another school conditional on having accepted the assignment. The independent variables are $\text{Dist}_{is}$, which is the distance between the assigned school $s$ and the student $i$, and $\mathcal{D}_{is}$, which captures differences in observable characteristics between assigned school $s$ and the previous school of student $i$. To compute the differences, we use information on school class sizes as measured by the student-to-teacher ratio, average parental education, average measure of academic performance (math),[26] and school size. We also include characteristics that reflect teaching quality, such as the proportion of novice teachers (with less than five years of experience). We use the prices 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 $\delta_i$. We include fixed effects per set of the same top 2 schools denoted by $\alpha_{dgl}$ to control for the systematic differences in the feasible set of schools, as well as an assigned school level fixed effect $\gamma_s$ that controls for the time invariant characteristics of the school. Finally, $e_{idsgl}$ represents the error term.

Below, we turn to specification (5) for estimating $\beta_1$, before exploring specification (7) in the following section to estimate $\psi$ and $\Lambda$ for the first stage, and assess variation in these estimates in the second stage. Finally, we examine the robustness of these results by applying less granular fixed effects, applying district $d$ by grade $g$ and education level $l$ fixed effects to control for the systematic differences in the feasible set of schools. Clustering of standard errors is done at the district level, further broken down by grade and education level for all presented regressions.

### 4.2. Main results

#### 4.2.1. The role of distance

Table 7 presents the results from specification (5),[27] with columns (1) and (3) showing estimates for whether parents accept or reject the assignment, as indicated by the dependent variable. In contrast,

[25] 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.

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

[27] To explore whether the effect of distance depends on the price of the school of origin, in Appendix A.2.4 we include an interaction term between these two variables in Eq. (5).

**Table 7**
The effect of distance on likelihood of accepting and staying at assigned school.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay |
|---|---|---|---|---|
| Distance (km) | −2.029*** | −3.649*** | −1.495*** | −1.967*** |
| | (0.508) | (0.626) | (0.126) | (0.248) |
| Price School of Origin | −0.015*** | −0.026*** | −0.014*** | −0.019*** |
| | (0.005) | (0.010) | (0.002) | (0.005) |
| Constant | 83.500*** | 84.771*** | 87.958*** | 82.814*** |
| | (1.501) | (3.053) | (0.529) | (1.150) |
| Observations | 12,698 | 7966 | 74,012 | 45,183 |
| R-squared | 0.239 | 0.301 | 0.108 | 0.175 |
| FE: Same Top 2 | X | X | | |
| FE: Female | X | X | X | X |
| FE: Dist-Gr-Lev | | | X | X |

(a) Note: In columns (1) and (2) we use same top 2 schools fixed effects and in columns (3) and (4) district-grade-education level fixed effects. Standard errors are clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

columns (2) and (4) showcase the data where the dependent variable reflects whether parents choose to stay or leave the assigned school. These results are conditional upon the initial acceptance of the assignment by the parents. Columns (1) and (2) rely on a unique fixed effect per set of the same top 2 schools, in contrast with columns (3) and (4), where we used district grade and education level fixed effects. The first set of fixed effects imposes a greater restriction on the estimation, as only those living very close to each other will have the same top two schools in their ranked based on proximity. This restriction allows us to directly mimic the randomization used by the DA algorithm. As mentioned, specification (5) 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). In contrast, columns (2) and (4) show the likelihood of remaining at the assigned schools conditional on having accepted their assignment.

The proximity of schools plays a crucial role in parental decisions, as seen in Table 7. Specifically, columns (1) and (3) show that for every additional kilometer a child must travel from home to school, the likelihood of parents accepting the assigned school decreases from 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 data reveals a more pronounced negative correlation between greater distances and the likelihood of students remaining at their initially assigned school. Notably, the negative impact intensifies in the second decision-making stage, increasing from 1.97 to 3.65 pp. This result is related to our theoretical model, which sheds light on the selection effect induced by the acceptance decision (see lemma 3.0.3). Parents who move to the public sector are more likely to value it more, and in particular be more sensitive to the particular characteristics that add value to it (like a shorter distance). This effect explains why the decision to remain at the public school is more strongly correlated to distance. This heightened sensitivity in the later stage suggests that parents may develop stronger apprehensions about longer distances as they progress through the decision-making process.

It is noteworthy that all schooling remained virtual at the time of these decisions. This suggests that parents were thinking long term, as in a virtual setting, distance is not a crucial factor. Arguably then, they saw the transition to public education as a more permanent solution as opposed to a short-term strategy.

A potential issue regarding the specification and causal interpretation of this model is whether the difference in distance between the closest and second-closest school is marginal. If so, the interpretation of the distance effect may lack substantive meaning. We address this concern in Appendix A.2.3. A second concern relates to siblings, who may affect both the causal interpretation and the randomness of the

**Table 8**
Non-linear effects of distance by sections.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay |
|---|---|---|---|---|
| Dist [1, 2] (km) | −1.884 | −7.065*** | −2.128*** | −3.676*** |
| | (1.294) | (1.985) | (0.409) | (0.647) |
| Dist [2, 3] (km) | −5.776*** | −7.165*** | −3.934*** | −4.880*** |
| | (1.475) | (1.912) | (0.414) | (0.699) |
| Dist [3, 4] (km) | −6.122*** | −11.949*** | −5.096*** | −7.276*** |
| | (1.871) | (2.658) | (0.477) | (0.879) |
| Dist [4, 5] (km) | −7.189*** | −14.479*** | −5.888*** | −7.675*** |
| | (2.167) | (2.750) | (0.554) | (1.001) |
| Price School of Origin | −0.015*** | −0.026*** | −0.014*** | −0.019*** |
| | (0.005) | (0.010) | (0.002) | (0.005) |
| Constant | 82.649*** | 83.828*** | 87.618*** | 82.596*** |
| | (1.450) | (3.000) | (0.569) | (1.107) |
| Observations | 12,697 | 7965 | 73,998 | 45,171 |
| R-squared | 0.239 | 0.301 | 0.109 | 0.175 |
| FE: Same Top 2 | X | X | | |
| FE: Female | X | X | X | X |
| FE: Dist-Gr-Lev | | | X | X |

(a) Note: Coefficients correspond to equation (5). In columns (1) and (2) we use same top 2 schools fixed effects and in columns (3) and (4) district-grade-education level fixed effects. Standard errors are clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

assignment process. In Appendix A.2.5, we conduct a robustness check in which we estimate the results using only students without siblings to assess the validity of our findings.

### 4.2.2. Non-linear effects: Distance

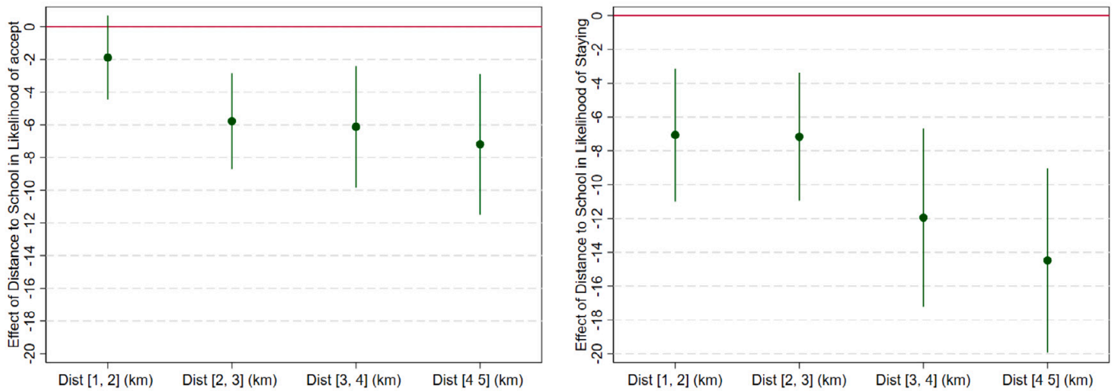
Table 8 presents the results of specification (6) and Fig. 6 displays the estimates for $\eta_r$ for $r \in \{2, 3, 4, 5\}$ from columns (1) and (2). Mirroring the format of the previous table, columns (1) and (3) present the results of the first stage, and columns (2) and (4) those of the second stage. In all specifications, we exclude the smallest distance category ($r = 1$). The coefficients are therefore meant to be interpreted as differences between each group relative to the first category (reference group).

Similarly, in Appendix A.2.8 we show the results for an alternative categorization, where quartiles separate the groups.

Fig. 6 (both panels) shows a linear relationship between distance from home to school and the likelihood of accepting and staying at the assigned school when allowing for non-linear effects in each categorized distance group. 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 group, vary between −1.88 p.p. up to −7.19 p.p. We cannot rule out that the effect sizes for these categories concerning the first group are the same, as we observe overlapping confidence intervals. The coefficients of the effect on the likelihood of remaining at the assigned school for each group, when compared to the smallest distance group, vary between −7.07 p.p. up to −14.48 p.p.

When students attend a school located more than 3 km away from their home, there is a 6 pp decrease in the likelihood that they remain at that assigned school, compared to when they are assigned to a school within 1 km of their home. This indicates that proximity is a significant factor for families when assessing the attractiveness of a school.

Fig. 6, 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 3 km. 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 from −7.07 in the [1,2) km range to −14.48 in the [4,5) km range. As detailed above, during the second stage, parents' decisions appear to be highly influenced by distance.




(a) On likelihood of Accepting

(b) On likelihood of Staying

**Fig. 6.** Non-linear effects of distance by sections.
Note: The figure presents the coefficients of regression Eq. (5). We use same top 2 schools fixed effects. Standard errors are clustered at the district by education level and grade. 95% confidence intervals are included.

#### *4.2.3. Heterogeneous effects by public school capacity and anticipated peer composition - Distance effect*

Motivated by concerns raised during the pandemic about potential changes in school composition — particularly the inflow of students from the private to the public sector — we examine whether these anticipated shifts influenced parental decision-making. As highlighted in previous literature, parents often place high value on a more advantaged socioeconomic peer environment. Therefore, the expectation of compositional changes may affect the likelihood of accepting an offer from the public system.

In this section, we analyze how observed peer composition prior to the assignment influences both the initial decision to accept and the subsequent decision to remain in the assigned school. To address this, we compare schools with higher and lower levels of spare capacity at baseline, using capacity as a proxy for the likelihood of changes in school composition. These estimates may be informative under the assumption that families can observe or reasonably infer the socioeconomic composition of schools beforehand. This information likely plays a role in shaping preferences and may underscore the relevance of peer composition in parental decision-making.

While we lack data on spare capacity by grade level, we do observe aggregate spare capacity by educational level within schools. We define high-capacity schools as those with total spare capacity above the median (550 seats), and low-capacity schools as those with capacity below this threshold.

The results showed in Table 9 suggest that in schools with low spare capacity — where peer composition is less likely to change significantly, but harder to anticipate — the probability of accepting an assignment is lower compared to high-capacity schools. Moreover, distance becomes less relevant when parents are assigned to low-capacity schools, suggesting that anticipated peer composition is a key factor in the decision-making process.

Our interpretation is as follows: in low-capacity schools, the student composition is expected to remain largely unchanged, limiting the entry of new students and thereby increasing uncertainty about whether any compositional change will occur. Conversely, high-capacity schools — given the structure of the assignment process — are more likely to admit new students from similar backgrounds or neighborhoods. This may reduce uncertainty and make these schools more attractive, particularly to families transitioning from the private sector who seeking continuity in peer environments.

In the second stage, we observe similar patterns, but the interpretation differs. In this selected sample, where families have already accepted their assignment, uncertainty is no longer present. Here, students assigned to schools with high spare capacity at baseline are more likely to remain enrolled, likely because the realized peer composition matched their initial expectations. In line with our theoretical model, students in this group appear more sensitive to distance than those in the initial decision stage.

In the context of our theoretical framework, this indicates that the potential value of a given public school ($V_{PU}$) is not only given by its distance to the family home, but also by its peer composition, a factor that has been well-established by the literature.

In brief, families are less likely to accept offers from schools where peer composition is perceived to be more uncertain and are also less likely to remain enrolled in schools where the peer composition turned out to be markedly different. This suggests that anticipated peer composition plays an important role in shaping both acceptance and retention decisions. However, we acknowledge that this analysis is descriptive and does not support causal inference; it is intended primarily to illustrate potential correlates of parental choice.

#### *4.2.4. The determinants of parents' preferences*

Table 10 presents specification (7). As in the previous table, columns (1) and (3) display the results for the first stage, while columns (2) and (4) present those for second stage, conditional on having accepted the assigned school in the first stage. However, due to the presence of unobserved variables that may be correlated with our observed variables, specification (7) 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 assigned school. These determinants are the ones that characterize, from the family's point of view, the attractiveness of the public sector option offered to them, $V_{PU}$ in our model. Our estimates in these specifications correspond to the sensitivity of parents' decisions to this value $V_{PU}$, and therefore correspond to $\frac{\partial \mathbb{P}(try)}{\partial V_{PU}}$ and $\frac{\partial \mathbb{P}(stay)}{\partial V_{PU}}$.

In Table 10, we consider the relative characteristics of the schools that students are assigned to compared to the characteristics that families experienced in their previous private schools. Like the previous section, columns (1) and (2) rely on same top 2 schools fixed effects and

**Table 9**
Heterogeneous effects of assignment acceptance by public school spare capacity.

| VARIABLES | Low spare capacity | | High spare capacity | |
|---|---|---|---|---|
| | (1) accept | (2) stay | (3) accept | (4) stay |
| Distance (km) | −1.484* | −2.634* | −2.538*** | −3.961*** |
| | (0.826) | (1.475) | (0.523) | (0.525) |
| Price School of Origin | −0.012* | −0.025 | −0.015*** | −0.028*** |
| | (0.006) | (0.015) | (0.005) | (0.009) |
| Constant | 80.092*** | 82.940*** | 86.704*** | 86.851*** |
| | (2.325) | (5.536) | (1.609) | (2.087) |
| Observations | 5052 | 3133 | 7054 | 4341 |
| R-squared | 0.318 | 0.390 | 0.266 | 0.329 |
| FE: Same Top 2 | X | X | X | X |
| FE: Female | X | X | X | X |

(a) Note: Coefficients correspond to equation (4). All specifications include fixed effects for students sharing the same top two preferred schools. Columns (1) and (2) restrict the sample to students assigned to schools with aggregate spare capacity above the median. Columns (3) and (4) focus on students assigned to schools with aggregate spare capacity below the median. Standard errors have been clustered at the district by level and grade level. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

**Table 10**
Relationship of school characteristics on likelihood of accepting and staying at assigned school.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay |
|---|---|---|---|---|
| Distance (km) | −1.732*** | −3.395*** | −1.206*** | −1.821*** |
| | (0.593) | (0.568) | (0.140) | (0.229) |
| Dif. Mean Math sc (s) | 7.662*** | 6.476** | 5.471*** | 4.084*** |
| | (1.792) | (2.533) | (0.643) | (0.780) |
| Dif. Mean Parent Edu. | 2.388*** | 6.093*** | 1.785*** | 3.896*** |
| | (0.844) | (1.549) | (0.374) | (0.419) |
| Log (Ratio Num St.) | 1.772 | 6.233*** | 2.259*** | 4.231*** |
| | (2.104) | (1.678) | (0.580) | (0.743) |
| Log (Ratio Num St.)$^2$ | 0.162 | −1.204** | −0.040 | −0.589*** |
| | (0.562) | (0.548) | (0.163) | (0.177) |
| Dif. STR | −0.042 | −0.075 | −0.093*** | −0.046 |
| | (0.093) | (0.126) | (0.034) | (0.046) |
| Dif. Prop. Novice Teachers | −1.229 | −0.462 | −1.587** | −1.449** |
| | (1.910) | (2.351) | (0.698) | (0.671) |
| Price School of Origin | −0.002 | −0.003 | −0.005*** | −0.007*** |
| | (0.003) | (0.009) | (0.001) | (0.003) |
| Constant | 78.330*** | 81.749*** | 83.346*** | 82.230*** |
| | (2.537) | (3.297) | (0.805) | (1.090) |
| Observations | 9476 | 5986 | 50,881 | 32,987 |
| R-squared | 0.263 | 0.337 | 0.117 | 0.176 |
| FE: Same Top 2 | X | X | | |
| FE: Female | X | X | X | X |
| FE: Dist-Gr-Lev | | | X | X |

(a) Note: Coefficients presented corresponds to equation (6). In columns (1) and (2) we use same top 2 schools fixed effects and in columns (3) and (4) district-grade-education level fixed effects. Standard errors are clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

columns (3) and (4) use district-grade-education level fixed effects. It is noteworthy that there is a significant decrease in the number of observations between Table 7 and column (1) of Table 10. The discrepancy occurs because we exclude clusters of observations that lack adequate variation in the acceptance rate at our chosen observation level (same top 2 schools). Additionally, some of the covariates in our analysis have missing values for some students.

As observed in the previous section, the estimated effect of distance remains consistent with the results in Table 7, thus highlighting the negative effect that distance has on both the likelihood of accepting as well as remaining at the assigned school. Beyond the role of distance, our data suggests that families care about school quality, which can be examined through mathematics assessment scores. We use the school's average standardized math score from 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, Kane, and Staiger (2006)). In particular, we find that an increase of one standard deviation in the difference in mean math scores between the assigned and origin school corresponds to an increase in the likelihood of accepting the assignment of between 5.47 and 7.66 p.p., as well as an increase in the likelihood of staying in the assigned school of between 4.1 and 6.5 p.p. The notable role of student performance observed in both the initial and subsequent stages underlines the importance placed on school quality. This result is consistent with findings reported by Beuermann, Jackson, Navarro-Sola, and Pardo (2023) and Abdulkadiroğlu et al. (2020), among others, who show that parents prefer schools of higher quality.

In addition to distance and school quality, families also value peer demographics. A unit increase in the average level of parental education,[28] (i.e., in academic degree category) between the assigned school and the private school of origin corresponds to an increase in the likelihood of accepting the assigned school of between 1.79 p.p. and 2.39 p.p. This suggests that peer demographics are a relevant factor in school choices. Further, the inclination to remain at the assigned school sees an increase on point estimates, with an associated rise in likelihood ranging from 3.9 to 6.1 percentage points. As previously discussed, parents in the origin private schools typically have higher education levels compared to those in the assigned public schools. This discrepancy suggests that these families may prefer to maintain positive peer demographics, and thus desire schools where the parental education level is consistent with their prior experiences. Moreover, this result is consistent with the heterogeneous effects by additional capacity discussed in the previous section. We can interpret these differences between the decision to try out the public sector and remain in it using our theoretical framework. The selection effect implies that parents that try out the public sector value its attributes more, and are therefore more sensitive to these characteristics (like peer demographics).

Parental education levels might, in addition, serve as an indirect measure of other valued school features not strictly related to peer demographics. For instance, parents may assume that schools with more educated parents are higher performing (possibly due to increased advocacy and pressure from these parents), safer, or have other qualities not directly related student performance. In the absence of test score data, peer demographics could also be used by parents as a proxy for student achievement (Elacqua, Schneider, & Buckley, 2006).

Relative school size is yet another factor that parents consider. A unit increase of the log ratio in school size — measured by the number of enrolled students — between the assigned and origin school is associated with an increase in the probability of accepting and staying at the assigned school. It may be that families interpret a significantly larger enrollment of the assigned school, compared to the origin school, as an indicator of the school's quality or popularity. However, the relationship is not linear.[29] The likelihood of accepting and staying is notably higher when the assigned school's size greatly exceeds that of the private school of origin.

The results also show that in the assignment stage, parents are less likely to accept being assigned to schools with larger student–teacher ratios (STR). However, this result does not hold for our main specification. Interestingly, this variable does not have a significant effect on the decision to remain in the assigned school. Analysis of another dimension of teacher quality — share of novice teachers — reveals that parents are less likely to accept and stay in an assigned school with a higher proportion of such teachers. Although this result

[28] As shown in Table 5 average parental education level is close to completed secondary education, with an average value of 4.9.

[29] We test for joint significance of the quadratic terms: in the first column, the F-statistic is 19.99, while in the second column, it is 17.17, i.e., both are jointly significant.

does not hold in our more stringent fixed effect specification, it suggests that families prefer schools with more experienced teachers.

Finally, we know that families who pay higher tuition at their private school of origin are less likely to accept the public school assignment. Tuition can be seen as an indicator of a family's purchasing power. Those with more financial resources and who thus have access to a broader range of schools tend to 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. Consequently, they might be less willing to transition their children away from what they perceive as higher-quality private institutions to public schools.

Our findings are consistent with evidence reported by Jacob and Lefgren (2007) and Burgess et al. (2015), who show that families prefer higher-performing schools with better peer demographics, 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) observe that while parents prefer schools with higher test scores and higher value-added, they also weigh non-academic factors, including the availability of after-school care, extracurricular activities, whether the child has a sibling in the school, and the school's proximity to home. Additionally, (Glazerman & Dotter, 2017) report that commuting distance, peer demographics, and academic indicators are essential aspects in school choice. They also highlight considerable heterogeneity of preferences by parent demographics.

In Table 11, we standardize the same variables detailed in Table 10. This approach allows us to discern the relative significance of each dimension in decisions to accept and remain at the assigned school. While many of the factors impacting the decision to accept the assigned school hold similar weight, school quality (evidenced by the difference in standardized average math scores) and school size appear to be the most important considerations. These are both aspects that parents are likely to be aware of before enrollment. However, once families have had first-hand experience with the school, other factors such as school proximity and average parental level of education become more influential in their decision to remain at the school. For instance, when the variables are standardized, the effect of difference in standardized average math score decreases from 4.3 pp. in the first stage to 3.63 in the second stage. Meanwhile, the effect of the difference in mean parental level of education increases from 2.52 pp. in the first stage to 6.43 in the second.

In contrast to specification (7), we could have used the observable characteristics of the assigned schools when evaluating the decisions to accept and remain enrolled. Table 31 in the Appendix presents the results of these specifications. There are evident similarities when comparing the results with those in Table 10, particularly concerning the influence of distance and peer quality. However, there are also some differences. Most notably, the quality of the assigned school does not appear to impact the decision to accept the school assignment or the decision to continue attending. Interestingly, school size only becomes a significant factor when deciding whether or not to stay at the assigned school.[30] We also find that the proportion of novice teachers is relevant and robust in the acceptance stage. This might suggest that younger teachers are viewed favorably, possibly 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.

[30] We test the joint significance in the first stage and find null results for both coefficients of school size (p-value = 0.32).

**Table 11**
Relationship of school characteristics on the likelihood of accepting and staying at assigned school - standardized.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay |
|---|---|---|---|---|
| Distance (km) | −3.463*** | −6.789*** | −2.411*** | −3.641*** |
| | (1.186) | (1.135) | (0.280) | (0.458) |
| Dif. Mean Math sc (s) | 4.298*** | 3.633** | 3.069*** | 2.291*** |
| | (1.005) | (1.421) | (0.360) | (0.438) |
| Dif. Mean Parent Edu. | 2.520*** | 6.431*** | 1.884*** | 4.112*** |
| | (0.890) | (1.635) | (0.395) | (0.442) |
| Log (Ratio Num St.) | 2.001 | 7.041*** | 2.552*** | 4.779*** |
| | (2.376) | (1.895) | (0.656) | (0.839) |
| Log (Ratio Num St.)$^2$ | 0.480 | −3.565** | −0.119 | −1.744*** |
| | (1.666) | (1.622) | (0.484) | (0.524) |
| Dif. STR | −0.398 | −0.707 | −0.882*** | −0.431 |
| | (0.878) | (1.194) | (0.319) | (0.434) |
| Dif. Prop. Novice Teachers | −0.463 | −0.174 | −0.599** | −0.547** |
| | (0.720) | (0.887) | (0.263) | (0.253) |
| Price school of origin | −0.344 | −0.541 | −0.993*** | −1.424*** |
| | (0.665) | (1.734) | (0.263) | (0.509) |
| Constant | 73.951*** | 69.688*** | 80.013*** | 75.158*** |
| | (0.391) | (0.574) | (0.084) | (0.165) |
| Observations | 9476 | 5986 | 50,881 | 32,987 |
| R-squared | 0.263 | 0.337 | 0.117 | 0.176 |
| FE: Same Top 2 | X | X | | |
| FE: Female | X | X | X | X |
| FE: Dist-Gr-Lev | | | X | X |

(a) Note: Coefficients presented corresponds to equation (6). In columns (1) and (2) we use same top 2 schools fixed effects and in columns (3) and (4) district-grade-education level fixed effects. Standard errors have been clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

#### 4.2.5. Heterogeneous effects by public school capacity and anticipated peer composition - Determinants of parents preferences

In this section, we examine whether parents' decisions to accept and remain in an assigned school vary according to the baseline spare capacity of public schools — that is, the number of vacant seats available prior to the assignment process. This classification captures the extent to which the peer composition of a school is expected to change following the entry of newly assigned students. In particular, schools with high spare capacity are likely to experience larger changes in peer composition, whereas schools with low spare capacity are expected to maintain a more stable student body. However, we assume that the informational environment differs: while peer composition in low-capacity schools is less likely to change, it is more difficult for families to anticipate; in contrast, although high-capacity schools will admit more new students, those students are likely to come from similar neighborhoods, given that distance was the key factor in the assignment algorithm. Thus, uncertainty is lower in high-capacity schools.

In Table 12, we can observe that in the first stage — when parents must decide whether to accept an assignment under noisy information — this distinction plays a crucial role. In line with our theoretical framework, the expectation of peer composition conditional on high spare capacity is higher than the expectation conditional on low spare capacity. This is because high-capacity schools are more likely to incorporate students from similar socioeconomic backgrounds or geographic areas, which parents can reasonably infer. As a result, parents assigned to high-capacity schools give more weight to other known attributes, such as academic performance and distance, when deciding whether to accept the offer.

In contrast, when assigned to low-capacity schools, parents form a lower expectation of peer composition. This implies that even if the assigned school has high academic performance, its perceived value is discounted due to the anticipated peer environment. Therefore, academic performance becomes less relevant in the decision to accept the assignment.

**Table 12**
Heterogeneous effects of assignment acceptance by public school spare capacity.

| VARIABLES | Low spare capacity | | High spare capacity | |
|---|---|---|---|---|
| | (1) accept | (2) stay | (3) accept | (4) stay |
| Distance (km) | −1.302 | −2.463* | −2.249*** | −3.547*** |
| | (0.855) | (1.476) | (0.719) | (0.649) |
| Dif. Mean Math sc (s) | 6.327*** | 3.184 | 9.932*** | 7.714** |
| | (2.263) | (5.002) | (2.803) | (3.249) |
| Dif. Mean Parent Edu. | 3.485*** | 7.414** | 1.122 | 4.482** |
| | (1.325) | (2.986) | (1.225) | (1.737) |
| Log (Ratio Num St.) | 1.845 | 7.814*** | −0.404 | 1.306 |
| | (3.180) | (2.818) | (3.079) | (2.959) |
| Log (Ratio Num St.)$^2$ | 0.865 | −2.045** | 0.261 | 0.073 |
| | (1.013) | (1.005) | (0.750) | (0.710) |
| Dif. STR | −0.009 | −0.086 | −0.003 | −0.096 |
| | (0.167) | (0.180) | (0.091) | (0.167) |
| Dif. Prop. Novice Teachers | 0.585 | −7.302* | −0.566 | 3.848 |
| | (2.501) | (4.239) | (2.749) | (2.817) |
| Price School of Origin | −0.000 | −0.001 | −0.004 | −0.009 |
| | (0.005) | (0.012) | (0.004) | (0.009) |
| Constant | 77.921*** | 81.972*** | 81.568*** | 86.293*** |
| | (3.232) | (6.488) | (4.241) | (3.522) |
| Observations | 4275 | 2506 | 4574 | 2957 |
| R-squared | 0.359 | 0.435 | 0.287 | 0.351 |
| FE: Same Top 2 | X | X | X | X |
| FE: Female | X | X | X | X |

()a) Note: Coefficients correspond to equation (6). All specifications include fixed effects for students sharing the same top two preferred schools. Columns (1) and (2) restrict the sample to students assigned to schools with aggregate spare capacity above the median. Columns (3) and (4) focus on students assigned to schools with aggregate spare capacity below the median. Standard errors have been clustered at the district by level and grade level. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In the second stage — when uncertainty is resolved and students decide whether to stay enrolled — a reversed pattern emerges. For students assigned to low-capacity schools, the realized peer composition may exceed their initially low expectations, leading to a higher probability of staying. The informational update has a positive effect. Conversely, academic performance becomes less influential in these schools, as it was already known and remains relatively stable.

In high-capacity schools, on the other hand, parents initially formed optimistic expectations about peer composition, which are more likely to be met. Therefore, the peer dimension becomes less salient at this stage. However, academic performance becomes more relevant in driving retention, likely because it is only fully revealed after experiencing the school environment, and because it may be more volatile in high-capacity schools due to the arrival of many new students.

Finally, as in the previous section, distance becomes less relevant when students are assigned to schools that are likely to retain the traditional demographic composition of the public sector, further supporting the idea that anticipated peer quality — shaped by baseline capacity and the structure of the assignment algorithm — is central to parental preferences.

These patterns suggest that spare capacity acts as a shock that alters the perceived value of public school offers, which is consistent with our theoretical framework where $VPU$ encapsulates all components of school utility, including peer composition.

## 5. Discussion

During the first year of the COVID-19 pandemic, Peru experienced an unprecedented rise in demand for public schools from students enrolled in private schools. More than one hundred thousand students requested a transfer. This surge may have been due to factors such as extended school closures, the inability of many private schools to move 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 induced by this system presents a unique opportunity to better understand the priorities of this specific group; namely, middle- and lower-middle class families who usually enroll their children in private schools (Balarin, Fontdevila, Marius, & Rodríguez, 2019). This setting furthermore offers potentially valuable insights for governments seeking to strengthen the supply of public schooling and reverse pre-pandemic trends in public and private education.

Our study sheds greater light on what these families value in a public school before and after directly experiencing the assigned school. We explore various school attributes such as distance, academic quality, and peer demographics in shaping their decisions to accept and remain in the public school system. While parents value all three of these attributes in both decision-making stages, we find that distance to school and peer demographics play a more significant role than academic quality in the decision to stay.

The results from this paper have important implications for policy.[31] First, controlling for other factors, parents care about student demographics, which is consistent with previous evidence. In the context of Latin America and the Caribbean, public schools tend to be less diverse and serve the most disadvantaged students. Therefore, attracting a more heterogeneous student body to the public sector is essential to foster diversity and bring in more middle-class families. Indeed, overly marked segregation means less diverse public schools, and potential stigmatization of this sector. Second, even after controlling for other relevant factors, a school's academic quality remains a priority in the decision to both accept and stay in the assigned school. In light of this finding, it is imperative that the Peruvian government continues to focus on school reforms that improve the quality of public schools. For instance, over the last decade, the government has instituted reforms to modernize the teacher career path (Elacqua et al., 2018) and attract and select more effective teachers into the profession.

Finally, a crucial policy insight is the role of experimentation in shaping parents' enrollment decisions. Specifically, short-term incentives allowing firsthand experiences with public services, like education, can have long-term consequences for a significant group of parents. Surprisingly, just 16.6%[32] of the students who experienced the assigned public school in 2020 returned to the private sector in 2021. Despite evidence that, all else equal, private schools do not have higher achievement than public schools (Balarin, 2015), parents in Peru still tend to view private schools as better.[33]. Publishing school-level information on test scores or value-added measures may help to reduce misconceptions about public and private school quality. The Peruvian government might also consider running campaigns to increase citizen knowledge about the objective attributes of public schools.

## CRediT authorship contribution statement

**Gregory Elacqua:** Conceptualization, Literature review, Funding, Methodology, Writing, Editing. **Nicolás Figueroa:** Conceptualization, Methodology, Formal analysis, Writing, Editing. **Andrés Fontaine:** Literature review, Data analysis, Methodology, Writing, Editing. **Juan**

[31] Care should, however, be taken in extrapolating these results since, in this setting, families experienced the public school imperfectly through online remote learning. Results might be different if schooling had been in-person.

[32] That is, only 9552 students out of the 57,496 who experienced at least a few months in the assigned public school in 2020 returned to the private sector after a year.

[33] Este País, 1 May 2014

**Margitic:** Conceptualization, Data analysis. **Carolina Méndez:** Conceptualization, Literature review, Funding, Methodology, Writing, Editing.

## Ethical considerations

This study did not involve human subjects, personal data, or animals requiring ethics approval.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.


## Acknowledgments

This work was supported by the Inter-American Development Bank, United States (Grant #RG-E1743)


## Appendix A

### A.1. Theoretical 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

$$\frac{\partial \mathbb{P}(try)}{\partial V_{PU}} = g \cdot \left[\mu_0 H_1'(V_{PU}) + (1-\mu_0) H_2'(V_{PU})\right] \tag{8}$$

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

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

$$\frac{\partial \mathbb{P}(stay)}{\partial V_{PU}} = g\left[\mu_1 \int_{\alpha=1}^{K} \alpha dF_1(\alpha) + (1-\mu_1)\int_{\alpha=1}^{K} \alpha dF_2(\alpha)\right] + \frac{\partial \mu_1}{\partial V_{PU}}\left[\int_{\alpha=1}^{K} G(L(\alpha V_{PU}))dF_1(\alpha) - \int_{\alpha=1}^{K} G(L(\alpha V_{PU}))dF_2(\alpha)\right] \tag{9}$$

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:

$$\underbrace{(\mu_1-\mu_0)(\bar{\alpha}_1-\bar{\alpha}_2) + \lambda(V_{PU})}_{\text{Selection effect}} - \underbrace{\left[(1-\mu_0)\int_{\alpha=\frac{V_{PR}^{(2)}}{V_{PU}}}^{\alpha=K} \alpha \cdot dF_2(\alpha) + \mu_0 \int_{\alpha=\frac{V_{PR}^{(2)}}{V_{PU}}}^{\alpha=K} \alpha \cdot dF_1(\alpha)\right]}_{\text{Horizon effect}} \quad \square$$

### A.2. Empirical appendix

#### A.2.1. Distribution of socio-economic level by school of origin

The following table presents the distribution of socio-economic index categories for the students' school of origin. We observe that more than half of the students come from schools classified as having a low or very low socio-economic level. (see Table 13).

**Table 13**
Distribution socio-economic level by school of origin.

| ISE Level | Private | Public |
|---|---|---|
| Very Low | 19,171 | 2798 |
| Low | 31,782 | 2124 |
| Middle | 30,465 | 1663 |
| High | 528 | 27 |
| Total | 81,946 | 6612 |

**Table 14**
Description round 1.

| Description | Total |
|---|---|
| Accept | 94,216 |
| Reject | 9960 |
| Change localization | 3240 |
| Invalid localization | 1938 |
| Pending | 15,886 |
| Total | 125,240 |

**Table 15**
Correlation address-related problem with various demographic characteristics and school-level features.

| VARIABLES | (1) Address problems | (2) Address problems |
|---|---|---|
| Mean Math sc (s) | 0.001 | 0.006 |
| | (0.009) | (0.004) |
| Mean Lang. sc (s) | 0.003 | −0.006 |
| | (0.009) | (0.004) |
| Mean Parent Edu. | −0.003 | −0.004*** |
| | (0.004) | (0.001) |
| Log (Num. Students) | −0.002 | 0.002 |
| | (0.002) | (0.001) |
| STR | −0.000 | −0.000 |
| | (0.000) | (0.000) |
| Prop. Novice Teachers | 0.002 | −0.000 |
| | (0.007) | (0.003) |
| Price School of Origin | −0.000*** | −0.000 |
| | (0.000) | (0.000) |
| Sample | Assigned school | School of Origin |
| Observations | 89,011 | 76,013 |
| R-squared | 0.088 | 0.092 |
| FE: Dist-Gr-Lev | X | X |
| FE: Female | X | X |

(a) Note: The coefficients correspond to regressions of a dummy variable indicating whether a student experienced an address-related problem on various demographic characteristics and school-level features. District-by-grade-education level fixed effects are included. Column (1) presents the results using characteristics of the assigned school, while Column (2) shows results using features of the school of origin. Standard errors are clustered at the district by grade and education level. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

#### A.2.2. Evidence for non-strategic manipulation of address

Here we present a description of the results from Round 1. As shown in Table 14, fewer than 5% of students experienced address-related issues. We construct a dummy variable indicating whether a student had an address-related problem and test its correlation with various demographic characteristics and school-level features. We find that no variables are significantly associated with address problems, except for average level of parental education in the school of origin. Interestingly, the correlation is negative — contrary to the concern that more educated or "strategic" parents might manipulate the system to gain access to preferred schools (see Table 15).

#### A.2.3. Robustness distance results

In the following tables, we present average distance to the assigned school by assigned proximity ranking, restricting the sample to students

**Table 16**
Regression and random Sample - Urban and Rural.

| Ranking | Urban | | Rural | |
|---|---|---|---|---|
| | Mean | Standard deviation | Mean | Standard deviation |
| 1 | 0.6858 | 0.4890 | 1.5407 | – |
| 2 | 1.2345 | 0.6482 | 2.4556 | – |
| 3 | 1.4189 | 0.6808 | 2.9350 | 0.0024 |
| 4 | 1.5623 | 0.6425 | 1.9720 | – |
| 5 | 1.7015 | 0.6115 | 2.0330 | – |
| 6 | 1.8656 | 0.6019 | 2.9746 | – |
| 7 | 1.9677 | 0.5849 | 1.6425 | – |
| 8 | 1.9900 | 0.5831 | 1.6712 | – |
| 9 | 2.1138 | 0.5525 | 2.9918 | – |
| 10 | 2.1159 | 0.5007 | – | – |
| 11 | 2.1971 | 0.4927 | – | – |
| 12 | 2.2402 | 0.5325 | – | – |
| 13 | 2.2747 | 0.4735 | – | – |
| 14 | 2.2997 | 0.4501 | – | – |
| 15 | 2.3010 | 0.4290 | – | – |
| 16 | 2.4064 | 0.3877 | – | – |
| 17 | 2.4865 | 0.3625 | – | – |
| 18 | 2.4584 | 0.3519 | – | – |
| 19 | 2.3243 | 0.3889 | – | – |
| 20 | 2.4118 | 0.3693 | – | – |
| 21 | 2.4227 | 0.3683 | – | – |
| Observations | 9239 | | 10 | |

(a) Note: The table reports the average distance (in kilometers) to the assigned school, disaggregated by the school's ranking according to proximity (e.g., first-closest, second-closest), and separately for urban and rural areas. The statistics are based on the sample that includes groups of students who share the same top two schools, where the first-choice school has full capacity.

**Table 17**
Sample - Urban and Rural.

| Ranking | Urban | | Rural | |
|---|---|---|---|---|
| | Mean | Standard deviation | Mean | Standard deviation |
| 1 | 0.6319 | 0.5222 | 0.9267 | 0.8835 |
| 2 | 1.0716 | 0.6027 | 2.0229 | 0.7658 |
| 3 | 1.2794 | 0.6563 | 2.1345 | 0.7653 |
| 4 | 1.4508 | 0.6231 | 2.3223 | 0.5044 |
| 5 | 1.5862 | 0.6198 | 2.2002 | 0.3139 |
| 6 | 1.7097 | 0.6013 | 2.4114 | 0.5919 |
| 7 | 1.7947 | 0.5820 | 2.2275 | 0.7115 |
| 8 | 1.8627 | 0.5894 | 1.6712 | – |
| 9 | 1.9723 | 0.5566 | 2.5024 | 0.4281 |
| 10 | 2.0387 | 0.5485 | 2.7057 | – |
| 11 | 2.1009 | 0.5267 | – | – |
| 12 | 2.1507 | 0.5136 | 1.7422 | – |
| 13 | 2.2148 | 0.4891 | – | – |
| 14 | 2.2439 | 0.5020 | 2.3748 | – |
| 15 | 2.2603 | 0.4680 | – | – |
| 16 | 2.2832 | 0.4522 | – | – |
| 17 | 2.3237 | 0.4508 | 2.9676 | – |
| 18 | 2.3529 | 0.4347 | – | – |
| 19 | 2.3472 | 0.4530 | – | – |
| 20 | 2.3713 | 0.4138 | – | – |
| 21 | 2.3653 | 0.4217 | – | – |
| Observations | 47,120 | | 422 | |

(a) Note: The table reports the average distance (in kilometers) to the assigned school, disaggregated by the school's ranking according to proximity (e.g., first-closest, second-closest), and separately for urban and rural areas. The statistics are based on the Regression Sample after applying all filtering steps.

**Table 18**
Interaction between distance and price of school of origin.

| VARIABLES | (1)<br>Accept | (2)<br>Stay | (3)<br>Accept | (4)<br>Stay |
|---|---|---|---|---|
| Distance (km) | −2.903*** | −5.482*** | −1.780*** | −1.774** |
| | (0.758) | (0.979) | (0.330) | (0.800) |
| Price School of Origin | −0.022*** | −0.046*** | −0.017*** | −0.018* |
| | (0.008) | (0.012) | (0.005) | (0.010) |
| Interaction | 0.004 | 0.008*** | 0.001 | −0.001 |
| | (0.002) | (0.003) | (0.002) | (0.003) |
| Constant | 85.408*** | 89.269*** | 88.573*** | 82.423*** |
| | (2.152) | (3.422) | (1.144) | (2.222) |
| Observations | 12,698 | 7966 | 74,012 | 45,183 |
| R-squared | 0.240 | 0.302 | 0.108 | 0.175 |
| FE: Same Top 2 | X | X | | |
| FE: Female | X | X | X | X |
| FE: District-Grade-Level | | | X | X |

(a) Note: In columns (1) and (2) we use same top 2 schools fixed effects and in columns (3) and (4) district-grade-education level fixed effects. Standard errors are clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

**Table 19**
Robustness check: Results using children without siblings.

| VARIABLES | (1)<br>accept | (2)<br>stay | (3)<br>accept | (4)<br>stay |
|---|---|---|---|---|
| Distance (km) | −2.238*** | −4.521*** | −1.800*** | −2.059*** |
| | (0.485) | (0.545) | (0.146) | (0.287) |
| Price School of Origin | −0.010** | −0.027** | −0.016*** | −0.024*** |
| | (0.005) | (0.012) | (0.002) | (0.004) |
| Constant | 82.848*** | 86.904*** | 88.561*** | 83.526*** |
| | (1.578) | (3.345) | (0.619) | (1.143) |
| Observations | 8118 | 4817 | 50,327 | 30,093 |
| R-squared | 0.316 | 0.381 | 0.132 | 0.209 |
| FE: Same Top 2 | X | X | | |
| FE: Female | X | X | X | X |
| FE: Dist-Gr-Lev | | | X | X |

(a) Note: Coefficients presented correspond to equation (5), estimated on a sample restricted to students without siblings in the school system at the time of assignment. In columns (1) and (2) we use same top 2 schools fixed effects and in columns (3) and (4) district-grade-education level fixed effects. Standard errors are clustered at the district by grade and education level. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

who share the same top two school choices. The results are further disaggregated by rural and urban areas.

Table 16 corresponds to the sample that includes groups of students who share the same top two schools, where the first-choice school has full capacity. Table 17 is based on the Regression Sample after applying all filtering steps.

Table 16 reports distances for a sample of students who share the same top two school options, where the first-choice school was at full capacity. Table 17 presents the same statistics for the regression sample after all filtering steps. In both cases, we observe clear and meaningful differences in average distance as the rank of the assigned school increases, indicating that these differences are not marginal.

### *A.2.4. Interaction between distance and price of school of origin*

To explore whether the effect of distance depends on the price of the school of origin, we include an interaction term between these two variables in Eq. (4). The rationale is that families coming from more expensive private schools might be more willing to accept a farther public school, given their potentially higher flexibility or different preferences.

We estimate the same specifications as in the main analysis, including both the fixed effects for the same top 2 schools and, alternatively, district-grade-level fixed effects.

**Table 20**
The effect of distance on likelihood of accepting and staying at assigned school.

| VARIABLES | (1)<br>accept | (2)<br>stay | (3)<br>accept | (4)<br>stay |
|---|---|---|---|---|
| Distance (km) | −1.311*** | −3.837*** | −1.163*** | −2.065*** |
| | (0.451) | (0.612) | (0.137) | (0.251) |
| Dif. Distance (km) | −0.714*** | 0.183 | −0.346*** | 0.111* |
| | (0.126) | (0.153) | (0.041) | (0.059) |
| Price School of Origin | −0.015*** | −0.025*** | −0.014*** | −0.019*** |
| | (0.005) | (0.010) | (0.002) | (0.005) |
| Constant | 82.122*** | 84.958*** | 87.224*** | 82.978*** |
| | (1.475) | (2.970) | (0.529) | (1.124) |
| Observations | 12,543 | 7872 | 72,788 | 44,488 |
| R-squared | 0.243 | 0.303 | 0.109 | 0.174 |
| FE: Same Top 2 | X | X | | |
| FE: Female | X | X | X | X |
| FE: Dist-Gr-Lev | | | X | X |

(a) Note: In columns (1) and (2) we use same top 2 schools fixed effects and in columns (3) and (4) district-grade-education level fixed effects. Standard errors are clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

**Table 21**
Relation distance and difference in covariates.

| VARIABLES | (1)<br>accept | (2)<br>stay | (3)<br>accept | (4)<br>stay |
|---|---|---|---|---|
| Dif. Mean Math sc (s) | −0.013 | 0.111 | −0.008 | −0.031 |
| | (0.073) | (0.076) | (0.039) | (0.049) |
| Dif Mean Lang sc (s) | 0.028 | −0.097 | 0.016 | 0.030 |
| | (0.071) | (0.069) | (0.039) | (0.049) |
| Dif. Mean Parent Edu. | −0.027 | −0.066** | 0.024 | 0.025 |
| | (0.021) | (0.030) | (0.018) | (0.019) |
| Log (Ratio Num St.) | 0.015 | 0.043 | 0.022 | 0.040** |
| | (0.030) | (0.049) | (0.016) | (0.019) |
| Log (Ratio Num St.)$^2$ | −0.002 | −0.012 | −0.000 | −0.007 |
| | (0.010) | (0.012) | (0.005) | (0.005) |
| Dif. STR | 0.001 | −0.001 | −0.003*** | −0.003** |
| | (0.001) | (0.003) | (0.001) | (0.001) |
| Dif. Prop. Novice Teachers | −0.007 | −0.082* | −0.002 | −0.014 |
| | (0.032) | (0.045) | (0.017) | (0.021) |
| Price School of Origin | 0.000 | 0.000 | 0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| Constant | 2.030*** | 1.929*** | 2.152*** | 2.121*** |
| | (0.049) | (0.057) | (0.033) | (0.035) |
| Observations | 9476 | 5986 | 50,881 | 32,987 |
| R-squared | 0.620 | 0.630 | 0.459 | 0.451 |
| FE: Dist-Gr-Lev | X | X | X | X |
| FE: Female | X | X | X | X |
| F-test: | 6.84 | 6.84 | 6.84 | 6.84 |
| Prob > F: | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

(a) Note: In columns (1) and (2) we use same top 2 schools fixed effects and in columns (3) and (4) district-grade-education level fixed effects. Standard errors are clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Results show that the main effects remain stable. The interaction term is only statistically significant in column (2), which corresponds to the second-stage decision (staying at the assigned school) under the fixed effects for the same top 2 schools. In this case, the positive and significant coefficient suggests that families from more expensive schools are indeed more likely to stay even if the assigned school is farther. In all other specifications, the interaction term is small and not significant (see Table 18).

### *A.2.5. Robustness siblings*

Siblings may affect both the causal interpretation of our estimates and the randomness of the school assignment process. For instance, if a student has a sibling already enrolled in a particular school, this may influence the likelihood of being assigned to that school, thus

**Table 22**
The effect of distance on likelihood of accepting and staying at assigned school - robustness check.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay | (5) accept | (6) stay |
|---|---|---|---|---|---|---|
| Distance (km) | −2.029*** | −3.649*** | −1.495*** | −1.967*** | −2.954*** | −4.323*** |
| | (0.508) | (0.626) | (0.126) | (0.248) | (0.582) | (0.991) |
| Price School of Origin | −0.015*** | −0.026*** | −0.014*** | −0.019*** | −0.023*** | −0.040*** |
| | (0.005) | (0.010) | (0.002) | (0.005) | (0.007) | (0.015) |
| Constant | 83.500*** | 84.771*** | 87.958*** | 82.814*** | 84.403*** | 86.919*** |
| | (1.501) | (3.053) | (0.529) | (1.150) | (2.298) | (4.513) |
| Observations | 12,698 | 7966 | 74,012 | 45,183 | 7204 | 4438 |
| R-squared | 0.239 | 0.301 | 0.108 | 0.175 | 0.292 | 0.338 |
| FE: Same Top 2 | X | X | | | | |
| FE: Female | X | X | X | X | X | X |
| FE: Same Top 3 | | | | | X | X |
| FE: Dist-Gr-Lev | | | X | X | | |

(a) Note: In columns (1) and (2) we use same top 2 schools fixed effects, in columns (3) and (4) district-grade-education level fixed effects and in columns (5) and (6) we use same top 3 schools fixed effects. Standard errors are clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

**Table 23**
The effect of distance on likelihood of accepting and staying at assigned school - robustness check.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay | (5) accept | (6) stay |
|---|---|---|---|---|---|---|
| Distance (km) | −1.311*** | −3.837*** | −1.163*** | −2.065*** | −2.124*** | −4.428*** |
| | (0.451) | (0.612) | (0.137) | (0.251) | (0.518) | (0.998) |
| Dif. Distance (km) | −0.714*** | 0.183 | −0.346*** | 0.111* | −0.900*** | 0.154 |
| | (0.126) | (0.153) | (0.041) | (0.059) | (0.206) | (0.214) |
| Price School of Origin | −0.015*** | −0.025*** | −0.014*** | −0.019*** | −0.025*** | −0.039*** |
| | (0.005) | (0.010) | (0.002) | (0.005) | (0.008) | (0.014) |
| Constant | 82.122*** | 84.958*** | 87.224*** | 82.978*** | 82.910*** | 87.082*** |
| | (1.475) | (2.970) | (0.529) | (1.124) | (2.365) | (4.451) |
| Observations | 12,543 | 7872 | 72,788 | 44,488 | 7121 | 4370 |
| R-squared | 0.243 | 0.303 | 0.109 | 0.174 | 0.295 | 0.340 |
| FE: Same Top 2 | X | X | | | | |
| FE: Female | X | X | X | X | X | X |
| FE: Same Top 3 | | | | | X | X |
| FE: Dist-Gr-Lev | | | X | X | | |

(a) Note: In columns (1) and (2) we use same top 2 schools fixed effects, in columns (3) and (4) district-grade-education level fixed effects and in columns (5) and (6) we use same top 3 schools fixed effects. Standard errors are clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

potentially violating the assumption of random assignment conditional on ranked schools based on proximity.

To address this concern, we conduct a robustness check in which we restrict the sample to students who do not have siblings in the school system at the time of assignment. This allows us to isolate the effects of the assignment process without the confounding influence of sibling-related preferences or allocation mechanisms.

The results of this analysis, presented in Table 19, are consistent with our main findings. In fact, the estimated coefficients are slightly larger in magnitude, suggesting that our baseline estimates may be conservative. This reinforces the validity of our identification strategy and supports the interpretation of the observed effects as causal.

*A.2.6. The role of difference in distance*

In this section, we enhance our primary specification (5), 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 having previously accepted the assignment.

$$y_{idsgl} = \alpha_{dgl} + \gamma_s + \delta_i + \beta_1 \cdot \text{Dist}_{is} + \beta_2 \cdot \text{Dif. Dist}_{is} + \beta_3 \cdot (\text{Price School Origin})_i + e_{idsgl} \tag{10}$$

Where $y_{idsgl}$ takes two values: first, $\text{Accept}_{idsgl}$, an indicator variable for student $i$ in district $d$, assigned school $s$, grade $g$ and education 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 she rejects her assignment; and second, $\text{Staying}_{idsgl}$, which takes a value of 100 if the student decides to stay at her assigned school in 2021 and a value of 0 if she moves to another school conditional on having accepted the assignment. The independent variables include $\text{Dist}_{is}$, the distance between the assigned school $s$ and student $i$'s address, and Dif. $\text{Dist}_{is}$, the difference in distance between the assigned school $s$ and student $i$'s address, compared to the distance between the school of origin and student $i$'s address. 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 $\delta_i$. We include district $d$ by grade $g$ and education level $l$ fixed effects denoted by $\alpha_{dgl}$ to control for the systematic differences in the feasible set of schools, as well as an assigned school level fixed effect $\gamma_s$, which controls for the time invariant characteristics of the school. Finally, $e_{idsgl}$ represents the error term.

**Table 24**
Non-linear effects of distance by quartiles.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay |
|---|---|---|---|---|
| Dist. [0.86, 1.90] (km) | −1.656 | −6.027*** | −2.394*** | −2.871*** |
| | (1.342) | (2.134) | (0.400) | (0.645) |
| Dist. [1.90, 3.25] (km) | −6.116*** | −7.946*** | −4.481*** | −5.007*** |
| | (1.636) | (2.048) | (0.443) | (0.766) |
| Dist. [3.25, 5] (km) | −5.516*** | −13.948*** | −5.520*** | −7.646*** |
| | (1.735) | (2.314) | (0.464) | (0.898) |
| Price School of Origin | −0.014*** | −0.026*** | −0.014*** | −0.019*** |
| | (0.005) | (0.010) | (0.002) | (0.005) |
| Constant | 82.497*** | 83.981*** | 87.891*** | 82.533*** |
| | (1.565) | (2.871) | (0.571) | (1.022) |
| Observations | 12,698 | 7966 | 74,012 | 45,183 |
| R-squared | 0.239 | 0.301 | 0.108 | 0.175 |
| FE: Same Top 2 | X | X | | |
| FE: Female | X | X | X | X |
| FE: Dist-Gr-Lev | | | X | X |

(a) Note: In columns (1) and (2) we use same top 2 schools fixed effects and in columns (3) and (4) district-grade-education level fixed effects. Standard errors are clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

**Table 25**
The effect of Distance on likelihood to accept and stay at assigned school — Primary Students.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay |
|---|---|---|---|---|
| Distance (km) | −1.529** | −3.932*** | −1.251*** | −1.967*** |
| | (0.597) | (0.714) | (0.168) | (0.270) |
| Price School of Origin | −0.011** | −0.021** | −0.016*** | −0.021*** |
| | (0.004) | (0.010) | (0.003) | (0.004) |
| Constant | 81.465*** | 84.884*** | 86.908*** | 83.691*** |
| | (1.639) | (3.259) | (0.707) | (1.220) |
| Observations | 7762 | 5356 | 39,862 | 27,216 |
| R-squared | 0.206 | 0.268 | 0.098 | 0.156 |
| FE: Same Top 2 | X | X | | |
| FE: Female | X | X | X | X |
| FE: Dist-Gr-Lev | | | X | X |

(a) Note: In columns (1) and (2) we use same top 2 schools fixed effects and in columns (3) and (4) district-grade-level fixed effects are used. Standard errors have been clustered at the district by level and grade level. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

**Table 26**
The effect of Distance on likelihood to accept and stay at assigned school — Secondary Students.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay |
|---|---|---|---|---|
| Distance (km) | −2.866** | −1.653 | −1.340*** | −1.245*** |
| | (1.129) | (1.194) | (0.327) | (0.359) |
| Price School of Origin | −0.032*** | −0.066*** | −0.019*** | −0.035*** |
| | (0.006) | (0.011) | (0.003) | (0.006) |
| Constant | 87.876*** | 93.141*** | 88.076*** | 89.871*** |
| | (2.913) | (3.358) | (1.221) | (1.711) |
| Observations | 3194 | 1836 | 18,628 | 11,648 |
| R-squared | 0.287 | 0.370 | 0.111 | 0.154 |
| FE: Same Top 2 | X | X | | |
| FE: Female | X | X | X | X |
| FE: Dist-Gr-Lev | | | X | X |

(a) Note: In columns (1) and (2) we use same top 2 schools fixed effects and in columns (3) and (4) district-grade-level fixed effects are used. Standard errors have been clustered at the district by level and grade level. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Interestingly, while 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 likelihood 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 for the decision 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%).

Table 21 further examines the potential independence between the assigned distance to school and other observable characteristics of the assigned school. Here, the dependent variable is the assigned distance to school, and the covariates are differences. The same fixed effect specifications from Eqs. (5) and (7) 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 is also likely uncorrelated with unobservable characteristics of the assigned schools.

### *A.2.7. The role of distance - Robustness check*

Tables 22 and 23 present a robustness check for specifications (5) and (6), respectively, where we have added columns (5) and (6) to Tables 7 and 20. For this analysis, we employ the same top 3 schools as fixed effects, thus imposing a stricter requirement for household proximity. We observe an increase in the magnitude of the effect of distance. Notably, there is a significant reduction in the number of observations from Fig. 1 to column (1). This decrease is attributed to the smaller group sizes when employing the same top 2 and same top 3 fixed effects, as there is no excess demand in most of these groups.

### *A.2.8. Non-linear effect: Distance - quartiles*

In this section, we adopt an alternative categorization of distance, as mentioned in Section 4.1.2. Specifically, we partition distance into quartiles based on the distribution of distance to school, represented by the following ranges: [0, 0.78), [0.78, 1.8), [1.8, 3.18), and [3.18, 5] (see Table Table 24).

Fig. 6 (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 group. We find that the coefficients of the effect on the likelihood of accepting the assigned school for each group, when compared to the smallest quartile of distance, vary between −2.12 p.p. up to −5.2 p.p. We cannot rule out that the effect sizes for all these categories concerning the first group are the same, as we observe overlapping confidence intervals. In addition, 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 −2.96 p.p. up to −7.58 p.p. (the effects can reach up to −13.41 when considering different sets of fixed effects). Finally, in contrast to panel (a), we observe that the effect sizes of the largest quartile group are statistically different when compared to that of the second quartile ([0.79,1.8)).

### *A.2.9. Differences by stage of education*

In this section we estimate the same regression of Table 7, but separately according to level of education, i.e., Primary and Secondary (see Tables 25 and 26, respectively).

Surprisingly, the effects differ considerably between both groups. Secondary students are more inclined to accept and remain in the assigned school than primary students.

Additionally, secondary students appear more sensitive to distance in the first stage. This may reflect their greater autonomy and daily experience with commuting. Conversely, this effect reverses in the second stage: primary students who initially accepted the assignment are more reluctant to remain in the assigned school when it is farther away. This suggests that families with younger children re-evaluate the burden of distance after gaining direct experience with the school.

**Table 27**
The effect of school characteristics on likelihood of accepting and staying at assigned school - robustness check.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay | (5) accept | (6) stay |
|---|---|---|---|---|---|---|
| Distance (km) | −1.732*** | −3.395*** | −1.206*** | −1.821*** | −2.961*** | −3.524*** |
| | (0.593) | (0.568) | (0.140) | (0.229) | (0.764) | (0.842) |
| Dif. Mean Math sc (s) | 7.662*** | 6.476** | 5.471*** | 4.084*** | 10.441*** | 9.101** |
| | (1.792) | (2.533) | (0.643) | (0.780) | (2.627) | (4.280) |
| Dif. Mean Parent Edu. | 2.388*** | 6.093*** | 1.785*** | 3.896*** | 0.656 | 6.599*** |
| | (0.844) | (1.549) | (0.374) | (0.419) | (1.543) | (2.131) |
| Log (Ratio Num St.) | 1.772 | 6.233*** | 2.259*** | 4.231*** | 2.867 | 5.852** |
| | (2.104) | (1.678) | (0.580) | (0.743) | (3.094) | (2.244) |
| Log (Ratio Num St.)$^2$ | 0.162 | −1.204** | −0.040 | −0.589*** | 0.261 | −1.137 |
| | (0.562) | (0.548) | (0.163) | (0.177) | (0.832) | (1.047) |
| Dif. STR | −0.042 | −0.075 | −0.093*** | −0.046 | −0.151 | −0.046 |
| | (0.093) | (0.126) | (0.034) | (0.046) | (0.107) | (0.158) |
| Dif. Prop. Novice Teachers | −1.229 | −0.462 | −1.587** | −1.449** | −1.167 | 0.597 |
| | (1.910) | (2.351) | (0.698) | (0.671) | (2.756) | (3.131) |
| Price School of Origin | −0.002 | −0.003 | −0.005*** | −0.007*** | −0.005 | −0.008 |
| | (0.003) | (0.009) | (0.001) | (0.003) | (0.007) | (0.012) |
| Constant | 78.330*** | 81.749*** | 83.346*** | 82.230*** | 74.388*** | 80.911*** |
| | (2.537) | (3.297) | (0.805) | (1.090) | (3.859) | (4.714) |
| Observations | 9476 | 5986 | 50,881 | 32,987 | 5424 | 3321 |
| R-squared | 0.263 | 0.337 | 0.117 | 0.176 | 0.317 | 0.386 |
| FE: Same Top 2 | X | X | | | | |
| FE: Female | X | X | X | X | X | X |
| FE: Same Top 3 | | | | | X | X |
| FE: Dist-Gr-Lev | | | X | X | | |

(a) Note: In columns (1) and (2) we use same top 2 schools fixed effects, in columns (3) and (4) district-grade-education level fixed effects and in columns (5) and (6) we use same top 3 schools fixed effects. Standard errors are clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In fact, although the effect of distance remains negative for secondary students in the second stage, it becomes statistically insignificant under our more demanding fixed effects specification, indicating that this group's persistence is less responsive to geographic constraints once the assignment has been accepted.

#### A.2.10. The determinants of parents' preferences - Robustness check

In this section, we conduct a robustness check of specification (7). The results are presented in Table 27, while those with the standardized variables are displayed in Table 28. The first four columns are the same as in Table 10. In columns (5) and (6) we change the set of fixed effects, using a unique fixed effect per set of the same top 3 schools. After applying the more stringent fixed effects, we do not observe significant differences in the second stage specification compared to that with the same top 2 schools fixed effect. However, there is a notable shift in both magnitude and statistical significance in the first stage. Notably, the number of observations between Fig. 1 and Table 27 drops, specifically from Fig. 1 to column (5).

#### A.2.11. Robustness results for determinants of parents

In this section, we estimate the same regression as in Table 10, but separately by education level (see Tables 29 and 30). Interestingly, the effects differ considerably between the two groups. The effect of distance shows similar patterns to those discussed in the previous section on the causal role of this factor. Academic performance appears to be more relevant for secondary students compared to primary students. Conversely, peer composition — measured by average parental level of education — is more important for primary students, especially in the first stage. In fact, at this stage, for secondary students, only academic performance and distance seem to be significant factors in school choice.

#### A.2.12. Alternative specification

In this final section, we explore an alternative specification similar to the main one. We also use two dependent variables: Accept$_{idsgl}$, an indicator for student $i$, district $d$, school $s$, education level $l$, and grade $g$ that takes a value of 100 if the student decides to accept her assignment in the first stage and a value of 0 if she rejects her assignment; and Staying$_{idsgl}$, which takes a value of 100 if the student decides to stay at her assigned school and a value of 0 if she moves to another school conditional on having accepted the assignment.

The independent variables of interest $\mathcal{X}_{sd}$ aim to capture the school $s$ at district $d$ characteristics of interest to parents when making their decision. First, we calculate the distance Dist$_{isd}$ between school $s$ and student $i$'s address in district $d$. We also include information on school class sizes (as measured by the student-to-teacher ratio), average parental level of education, average measures of academic performance (math), and school size. The additional inclusion of the proportion of novice teachers (with less than five years of experience) give us an idea of the relevance of teaching quality. We include the origin school's cost to capture an approximate measure of income level and a dummy to control for the student's gender denoted by $\delta_i$. District $d$ by grade $g$ and education level $l$ fixed effects denoted by $\alpha_{dgl}$ control for the systematic differences in the feasible offer sets of schools. Finally, $e_{isdgl}$ represents the error term.

In summary, we use the same variables as in specification (7), but in education levels instead of the differences between the assigned and origin schools. The results are presented in Tables 31 and 32.

**Table 28**
The effect of school characteristics on likelihood of accepting and staying at assigned school - robustness check - standardized.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay | (5) accept | (6) stay |
|---|---|---|---|---|---|---|
| Distance (km) | −3.463*** | −6.789*** | −2.411*** | −3.641*** | −5.921*** | −7.595*** |
| | (1.186) | (1.135) | (0.280) | (0.458) | (1.529) | (1.760) |
| Dif. Mean Math sc (s) | 4.298*** | 3.633** | 3.069*** | 2.291*** | 5.857*** | 4.306* |
| | (1.005) | (1.421) | (0.360) | (0.438) | (1.473) | (2.382) |
| Dif. Mean Parent Edu. | 2.520*** | 6.431*** | 1.884*** | 4.112*** | 0.692 | 7.519*** |
| | (0.890) | (1.635) | (0.395) | (0.442) | (1.628) | (2.031) |
| Log (Ratio Num St.) | 2.001 | 7.041*** | 2.552*** | 4.779*** | 3.239 | 9.254*** |
| | (2.376) | (1.895) | (0.656) | (0.839) | (3.495) | (2.578) |
| Log (Ratio Num St.)$^2$ | 0.480 | −3.565** | −0.119 | −1.744*** | 0.773 | −5.295* |
| | (1.666) | (1.622) | (0.484) | (0.524) | (2.465) | (2.960) |
| Dif. STR | −0.398 | −0.707 | −0.882*** | −0.431 | −1.421 | −0.754 |
| | (0.878) | (1.194) | (0.319) | (0.434) | (1.010) | (1.524) |
| Dif. Prop. Novice Teachers | −0.463 | −0.174 | −0.599** | −0.547** | −0.440 | −0.394 |
| | (0.720) | (0.887) | (0.263) | (0.253) | (1.040) | (1.146) |
| Price school of origin | −0.344 | −0.541 | −0.993*** | −1.424*** | −0.966 | −0.608 |
| | (0.665) | (1.734) | (0.263) | (0.509) | (1.490) | (2.606) |
| Constant | 73.951*** | 69.688*** | 80.013*** | 75.158*** | 70.367*** | 65.504*** |
| | (0.391) | (0.574) | (0.084) | (0.165) | (0.686) | (0.965) |
| Observations | 9476 | 5986 | 50,881 | 32,987 | 5424 | 3569 |
| R-squared | 0.263 | 0.337 | 0.117 | 0.176 | 0.317 | 0.370 |
| FE: Same Top 2 | X | X | | | | |
| FE: Female | X | X | X | X | X | X |
| FE: Same Top 3 | | | | | X | X |
| FE: Dist-Gr-Lev | | | X | X | | |

(a) Note: In columns (1) and (2) we use same top 2 schools fixed effects, in columns (3) and (4) district-grade-education level fixed effects and in columns (5) and (6) we use same top 3 schools fixed effects. Standard errors are clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

**Table 29**
The relationship of school characteristics on likelihood of accepting and staying at assigned school — Primary Students.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay |
|---|---|---|---|---|
| Distance (km) | −1.357** | −3.608*** | −1.214*** | −1.906*** |
| | (0.634) | (0.639) | (0.178) | (0.258) |
| Dif. Mean Math sc (s) | 5.139** | 4.172 | 4.447*** | 2.862*** |
| | (2.339) | (2.908) | (0.785) | (0.949) |
| Dif. Mean Parent Edu. | 2.808*** | 6.145*** | 1.690*** | 3.613*** |
| | (0.981) | (1.620) | (0.425) | (0.505) |
| Log (Ratio Num St.) | 4.919 | 5.906*** | 4.351*** | 4.261*** |
| | (3.003) | (1.972) | (0.793) | (1.085) |
| Log (Ratio Num St.)$^2$ | −0.336 | −1.147 | −0.475** | −0.477 |
| | (0.735) | (0.759) | (0.234) | (0.300) |
| Dif. STR | −0.090 | −0.184 | −0.086*** | −0.111* |
| | (0.097) | (0.152) | (0.033) | (0.060) |
| Dif. Prop. Novice Teachers | −2.271 | −0.356 | −1.651 | −1.517* |
| | (2.659) | (2.585) | (1.007) | (0.827) |
| Price School of Origin | 0.003 | −0.001 | −0.004** | −0.007** |
| | (0.003) | (0.008) | (0.002) | (0.003) |
| Constant | 74.073*** | 82.535*** | 80.657*** | 80.442*** |
| | (3.656) | (3.003) | (0.914) | (1.357) |
| Observations | 6652 | 4431 | 34,392 | 22,913 |
| R-squared | 0.244 | 0.318 | 0.113 | 0.173 |
| FE: Same Top 2 | X | X | | |
| FE: Female | X | X | X | X |
| FE: Dist-Gr-Lev | | | X | X |

(a) Note: Presented coefficients correspond to equation (6). In columns (1) and (2) we use same top 2 schools fixed effects and in columns (3) and (4) district-grade-education level fixed effects. Standard errors are clustered at the district by grade and education level. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

**Table 30**
The relationship of school characteristics on likelihood of accepting and staying at assigned school — Secondary Students.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay |
|---|---|---|---|---|
| Distance (km) | −2.599** | −2.277 | −1.019*** | −1.552*** |
| | (1.287) | (1.428) | (0.311) | (0.344) |
| Dif. Mean Math sc (s) | 14.251*** | 15.630*** | 9.120*** | 8.028*** |
| | (3.425) | (4.532) | (1.121) | (1.528) |
| Dif. Mean Parent Edu. | −0.144 | 6.006** | 1.501** | 4.438*** |
| | (1.826) | (2.936) | (0.626) | (0.793) |
| Log (Ratio Num St.) | −3.101 | 5.466* | −0.469 | 3.465*** |
| | (2.819) | (3.184) | (0.831) | (1.192) |
| Log (Ratio Num St.)$^2$ | 0.678 | −1.038 | 0.332* | −0.666** |
| | (0.721) | (0.950) | (0.191) | (0.280) |
| Dif. STR | 0.090 | 0.160 | −0.091 | 0.055 |
| | (0.183) | (0.175) | (0.058) | (0.059) |
| Dif. Prop. Novice Teachers | 1.824 | −1.575 | −0.979 | −1.802 |
| | (3.717) | (4.869) | (1.190) | (1.219) |
| Price School of Origin | −0.021*** | −0.007 | −0.007*** | −0.008** |
| | (0.006) | (0.015) | (0.002) | (0.004) |
| Constant | 90.295*** | 85.106*** | 88.359*** | 88.609*** |
| | (4.075) | (6.174) | (1.604) | (1.708) |
| Observations | 2749 | 1555 | 16,076 | 10,074 |
| R-squared | 0.315 | 0.406 | 0.130 | 0.181 |
| FE: Same Top 2 | X | X | | |
| FE: Female | X | X | X | X |
| FE: Dist-Gr-Lev | | | X | X |

(a) Note: Presented coefficients correspond to equation (6). In columns (1) and (2) we use same top 2 schools fixed effects and in columns (3) and (4) district-grade-education level fixed effects. Standard errors are clustered at the district by grade and education level. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A notable finding when comparing the results with Table 10 is that the mean math scores in the assigned schools have no significant effect in education levels, but do have a significant effect in differences. This seems to support the idea that parents base their decisions on the characteristics of the origin school.

**Table 31**
Alternative results - education levels.

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay | (5) accept | (6) stay |
|---|---|---|---|---|---|---|
| Distance (km) | −2.227*** | −3.100*** | −1.507*** | −2.044*** | −2.808*** | −3.171*** |
| | (0.453) | (0.462) | (0.141) | (0.224) | (0.560) | (0.640) |
| Mean Math sc (s) | −1.600 | 7.090** | 0.563 | 2.334 | 0.654 | 8.206** |
| | (2.541) | (2.917) | (0.777) | (1.549) | (3.625) | (4.071) |
| Mean Parent Edu. | 2.192* | 6.327*** | 1.396*** | 2.378*** | 1.908 | 7.244*** |
| | (1.259) | (1.672) | (0.495) | (0.826) | (1.911) | (2.669) |
| Log (Num Students) | −7.548 | 25.445** | 4.991 | 20.501*** | 15.115 | 54.064*** |
| | (16.571) | (12.585) | (4.834) | (7.084) | (20.015) | (19.025) |
| Log (Num Students)$^2$ | 0.669 | −1.962** | −0.355 | −1.368** | −1.308 | −4.188*** |
| | (1.318) | (0.957) | (0.389) | (0.556) | (1.576) | (1.419) |
| STR | 0.027 | 0.227 | −0.055 | 0.195** | 0.058 | 0.028 |
| | (0.173) | (0.209) | (0.051) | (0.090) | (0.247) | (0.295) |
| Prop. Novice Teachers | 8.028*** | 7.171* | 1.756 | −1.411 | 12.118*** | 7.740 |
| | (2.131) | (3.798) | (1.245) | (1.612) | (2.875) | (4.917) |
| Price School of Origin | −0.015*** | −0.029*** | −0.017*** | −0.025*** | −0.023*** | −0.042*** |
| | (0.004) | (0.008) | (0.002) | (0.004) | (0.006) | (0.014) |
| Constant | 91.349* | −34.905 | 64.674*** | −4.393 | 27.876 | −125.365** |
| | (51.020) | (40.152) | (14.791) | (23.138) | (62.816) | (62.626) |
| Observations | 11,448 | 7602 | 58,857 | 38,951 | 6818 | 4480 |
| R-squared | 0.118 | 0.136 | 0.051 | 0.063 | 0.130 | 0.149 |
| FE: Same Top 2 | X | X | | | | |
| FE: Female | X | X | X | X | X | X |
| FE: Same Top 3 | | | | | X | X |
| FE: Dist-Gr-Lev | | | X | X | | |

(a) Note: In columns (1) and (2) we use same top 2 schools fixed effects, in columns (3) and (4) district-grade-education level fixed effects and in columns (5) and (6) we use same top 3 schools fixed effects. Standard errors are clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

**Table 32**
Alternative results - education levels (standardized).

| VARIABLES | (1) accept | (2) stay | (3) accept | (4) stay | (5) accept | (6) stay |
|---|---|---|---|---|---|---|
| Distance (km) | −4.454*** | −6.199*** | −3.013*** | −4.086*** | −5.614*** | −6.340*** |
| | (0.907) | (0.923) | (0.282) | (0.447) | (1.119) | (1.280) |
| Mean Math sc (s) | −0.534 | 2.366** | 0.188 | 0.779 | 0.218 | 2.738** |
| | (0.848) | (0.973) | (0.259) | (0.517) | (1.210) | (1.358) |
| Mean Parent Edu. | 1.457* | 4.205*** | 0.928*** | 1.581*** | 1.268 | 4.814*** |
| | (0.837) | (1.111) | (0.329) | (0.549) | (1.270) | (1.774) |
| Log (Num Students) | −6.369 | 21.472** | 4.212 | 17.301*** | 12.755 | 45.624*** |
| | (13.984) | (10.620) | (4.079) | (5.978) | (16.891) | (16.055) |
| Log (Num Students)$^2$ | 6.418 | −18.831** | −3.407 | −13.129** | −12.560 | −40.208*** |
| | (12.651) | (9.191) | (3.733) | (5.333) | (15.129) | (13.619) |
| STR | 0.148 | 1.246 | −0.304 | 1.071** | 0.319 | 0.155 |
| | (0.951) | (1.150) | (0.282) | (0.495) | (1.359) | (1.621) |
| Prop. Novice Teachers | 1.407*** | 1.257* | 0.308 | −0.247 | 2.124*** | 1.357 |
| | (0.374) | (0.666) | (0.218) | (0.283) | (0.504) | (0.862) |
| Price school of origin | −3.080*** | −5.750*** | −3.348*** | −5.013*** | −4.648*** | −8.456*** |
| | (0.795) | (1.688) | (0.490) | (0.829) | (1.176) | (2.698) |
| Constant | 75.232*** | 70.102*** | 80.891*** | 74.447*** | 72.332*** | 66.882*** |
| | (0.880) | (0.868) | (0.275) | (0.391) | (1.124) | (1.544) |
| Observations | 11,448 | 7602 | 58,857 | 38,951 | 6818 | 4480 |
| R-squared | 0.118 | 0.136 | 0.051 | 0.063 | 0.130 | 0.149 |
| FE: Same Top 2 | X | X | | | | |
| FE: Female | X | X | X | X | X | X |
| FE: Same Top 3 | | | | | X | X |
| FE: Dist-Gr-Lev | | | X | X | | |

(a) Note: In columns (1) and (2) we use same top 2 schools fixed effects, in columns (3) and (4) district-grade-education level fixed effects and in columns (5) and (6) we use same top 3 schools fixed effects. Standard errors are clustered at the district by education level and grade. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The impact of Parent Education in education levels does not change significantly when compared to its effect in differences. Notably, in the second stage, school size has a substantial effect, which is absent in the first stage. Conversely, the relative size of the school has significant positive effects in both stages.

$$y_{isdgl} = \alpha_{dgl} + \delta_i + \eta \cdot \mathrm{Dist}_{isd} + \mathcal{X}'_{sd}\Omega + e_{isdgl} \quad (11)$$

## References

Abdulkadiroğlu, A., Agarwal, N., & Pathak, P. A. (2017). The welfare effects of coordinated assignment: Evidence from the new york city high school match. *American Economic Review*, *107*(12), 3635–3689.

Abdulkadiroğlu, A., Pathak, P. A., Schellenberg, J., & Walters, C. R. (2020). Do parents value school effectiveness? *American Economic Review*, *110*(5), 1502–1539.

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

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

Balarin, M. (2015). *The default privatization of Peruvian education and the rise of low-fee private schools: Better or worse opportunities for the poor?*. Privatisation in Education Research Initiative.

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

Balarin, M., Kitmang, J., Ñopo, H., & Rodríguez, M. F. (2018). *Mercado privado, consecuencias públicas: los servicios educativos de provisión privada en el Perú*: *Technical Report 89*, GRADE, URL https://www.grade.org.pe/wp-content/uploads/GRADEdi89.pdf.

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

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

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

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

Elacqua, G., Méndez, C., & Navarro, M. (2022). Colegios privados en tiempos de covid-19 en am00E9rica latina y El Caribe. http://dx.doi.org/10.18235/0004434, IADB Technical Note.

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

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

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

Hastings, J. S., Kane, T. J., & Staiger, D. O. (2005). *Parental preferences and school competition: Evidence from a public school choice program*. National Bureau of Economic Research Cambridge, Mass., USA.

Hastings, J. S., Kane, T. J., & Staiger, D. O. (2006). *Preferences and heterogeneous treatment effects in a public school choice lottery*. National Bureau of Economic Research Cambridge, Mass., USA.

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

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

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

Kleitz, B., Weiher, G. R., Tedin, K., & Matland, R. (2000). Choice, charter schools, and household preferences. *Social Science Quarterly*, 846–854.

Lavado, P. (2021). *Impacto del COVID-19 en el Empleo en el Perú: una aproximación por tipo de contrato y tamaño de empresa*: *Documento de trabajo*, CIUP - Centro de Investigación de la Universidad del Pacífico, URL https://ciup.up.edu.pe/repositorio/handle/20.500.12866/795.

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

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

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