The domino effect in centralized school assignment: the case of Mexico*

Arturo Aguilar† — *ITAM*

Adrián Martínez‡ — *ITAM and Banco de Mexico*

Cristián Sánchez § — *ITAM*

May 7, 2025

Abstract

This paper studies the consequences of a human error during the school choice mechanism of Mexico City public high school system. Using the reported rank-ordered lists of the applicants, we estimate their indirect utilities and generate cardinal measures of the error's impact on the applicants' utilities. In particular, we found that the median welfare change is negatively correlated with the educations of the applicants' mothers (proxy for socioeconomic level).

*The views and conclusions in this paper are the authors' exclusively and do not necessarily reflect those of Banco de México and its Board of Governors.

†ITAM, Av. Camino a Santa Teresa 930, Ciudad de México 10700, México (e-mail: arturo.aguilar@itam.mx).

‡Banco de México, (e-mail: adrian.martinez@banxico.org.mx)

§ITAM, Av. Camino a Santa Teresa 930, Ciudad de México 10700, México (e-mail: cristian.sanchez@itam.mx).

1 Introduction

Around the world, different cities use centralized assignment algorithms to match schools and students. Economists have been studying how different assignment algorithms affect efficiency in matching students and schools. Abdulkadiroğlu and Sönmez (2003) were the first to approach school choice as a mechanism design problem and found that Boston, Colombus, Minneapolis, and Seattle's school assignment mechanisms had serious efficiency problems. For example, the school programs of these cities had procedures where students could lose their priorities in some schools if they did not list that school as their favorite option, which generated "very complicated admissions games" and "inefficient allocation of school seats" (Abdulkadiroğlu and Sönmez, 2003). Since then, the development of the literature on assignment and matching algorithms has contributed to increasing the "transparency, efficiency, and equitability" of the centralized assignment mechanisms (Elacqua et al., 2021).

This paper studies the implications of an error during the 2017 school choice mechanism of the Metropolitan Area of Mexico City. The Mexico City public school choice mechanism has around 300,000 applicants and 700 school programs per year, and around 216,000 and 253,000 students were enrolled in 2007 and 2017. In comparison, the New York City public school district accepts around 90,000 ninth graders per year (Abdulkadiroğlu et al., 2020). Due to the large scale of the Mexico City school choice mechanism, analyzing it and its outcomes could generate insights on how to improve the welfare of many high school students.

To our knowledge, this is the first paper to analyze the error in Mexico's school choice mechanism and its implications in a centralized school assignment system. Mexico City's centralized system uses the Serial Dictatorship algorithm, ranking students by the admission exam score. Indeed, the error affected the admission exam scores of approximately 14,000 students due to a human mistake while grading the exams. The authorities noticed the error after the publication of the results; thus, they decided to maintain the original assignments of those students with correct exam scores and changed the assignments of the students with incorrect exam scores.

The error in exam scores affected only 14,000 students, but the students with correct exam scores were also affected indirectly. Changing one of the algorithm's inputs (the students' exam scores and rank-ordered lists) makes it pick a different final assignment. In the case of the 2017 error, we show in the next sections that the final assignments of the 2017 contest, with the correction made by the authorities, are not similar to the match outcomes of an algorithm that uses the correct exam scores. We estimate that between 12,349 and 74,767 students got a different assignment compared to a "universe" without error. The latter is a result of the spillover effects generated by the error. If student i has a score equal to 100, the algorithm will try to find a spot in a school of her rank-ordered list given that it already assigned students with scores greater than 100. Therefore, if the algorithm's priority index (exam score) is changed, it will generate different assignments. Thus, even if a student did not receive an incorrect score, the change in the exam scores of some students could impact her final assignment.

To analyze the impact of the error in the final match outcomes and, therefore, on welfare, we simulate the match outcomes that should have happened if the error had never occurred. Because we have the participants' rank-order lists, we use a standard rank-order logit specification to compute the expected utility of student i when attending school j. Using the simulated assignments and the estimated expected utilities, we compute the welfare changes caused by the error. In particular, we find that the error had a median positive impact on welfare equivalent to decreasing the distance between the students' houses and schools by 70 kilometers. We also find that the median welfare change is negatively correlated with the student's mother's education (a proxy for socioeconomic level). The latter result could imply that the most vulnerable students benefited most from the error.

The error affected the index priority of several students with particular characteristics. As shown in the following sections, the students affected by an incorrect exam score had higher average GPA, admission exam scores, and parental education. In simple terms, the reduction in the priority index of these students allowed other students with lower exam scores to be admitted to educational options that, without the error, would have had higher exam score cut-offs. Since we show that the welfare consequences of the error are negatively correlated with the applicants' mother's education level (a proxy for socioeconomic level), we can interpret this result as that the current exam score priority mechanism benefits applicants of higher socioeconomic levels by giving them access to higher quality (highly demanded) educational options. This motivates the discussion of whether the current priority structure could be improved to give access to higher-achieving schools to the most vulnerable groups of applicants.

This paper is related to studies that investigate the consequences of the current school choice mechanism of Mexico City's public high school system and offers possible changes to increase the participation of vulnerable groups in higher-achieving schools. For instance, Pariguana and Ortega-Hesles (2022) shows that adding other skills measures such as GPA to the priority structure of Mexico

City's school admission system increases the quantity of female and lower-income students that would be admitted to elite schools (*IPN* and *UNAM*). On the other hand, Ngo and Dustan (2022) studies the gender gaps in science, technology, engineering, and mathematics (STEM) during high school in Mexico City and finds that changes to the assignment priority structure and preference-altering interventions could decrease the STEAM gap depending on the students' achievements.

2 High-School Education in Mexico City

2.1 Mexico City School Choice Mechanism

In 1996, the educational institutions that provide public high school education in the Metropolitan Area of Mexico City (MAMC) signed an agreement to initiate a centralized assignment mechanism. The process consists of several stages in which students compete for a seat in public high schools of MAMC.[1] The institution created to coordinate the centralized matching mechanism is called *COMIPEMS* (*Comisión Metropolitana de Instituciones Públicas de Educación Media Superior*). This institution centrally manages the assignment of around 300,000 students per year to the different public institutions –with independent administrations– that comprise the MAMC's high school system.

The schools that participate in the assignment process belong to one of 15 sub-systems.[2] Before its creation, each sub-system had its own admission procedures, which complicated the application process for the students and possibly generated inefficient matching outcomes (Manjunath and Turhan, 2016). Each sub-system has a particular curriculum, with the *UNAM* and *IPN* sub-systems being the most popular.[3] With the creation of COMIPEMS, the application and admission process became homogeneous across all sub-systems.[4] *Table 1* includes descriptive statistics for the subsystems: number of programs/schools, the number of students that recorded a program in each of the sub-systems as

[1] This area encompasses Mexico City and 22 surrounding municipalities

[2] A sub-system is a set of schools administered by a certain entity. For example, the *UNAM* sub-system is administered by one of Mexico's most prestigious public universities, which is autonomous from the government. Other sub-systems are administered by federal/state ministries. They also differ in the type of curriculum. For example, the *IPN* sub-system is well-known for its curriculum, which specializes in engineering and mathematics.

[3] The schools affiliated with these sub-systems are considered the elite high schools. They are respectively affiliated to two highly prestigious public universities: (*Universidad Nacional Autónoma de México* (UNAM) and *Instituto Politénico Nacional* (IPN) (Pariguana and Ortega-Hesles, 2022). The *UNAM* subsystem has one particular benefit for its students: after completing their high school studies, they receive automatic placement at *UNAM* university without taking the admission exam, which is quite competitive.

[4] The only exception to this rule is that UNAM and IPN establish a minimum GPA of 7 out of 10 at middle school. During the 2017 admission process 18,299 students applied to UNAM or IPN without having the required GPA.

their first option, the percentage of the rank-ordered lists that mention at least one program of each sub-system, and the type of education offered by each sub-system. Hereon, we define an educational option as a combination of a school and a program. This implies that we treat two programs (e.g., engineering and administration) as different educational options available at the same high school.

Table 1: Sub-systems: summary statistics

Sub-system	Number of schools	First option	In preferences	Type of Education
Colegio Bachilleres	20	9,759	49.6%	Bachillerato General
CONALEP - CDMX	89	6,890	25.79%	Profesional Técnico
D.G Bachillerato	2	708	4.42%	Bachillerato General
Agropecuaria	3	486	1.72%	Bachillerato Tecnológico
Industrial	51	16,974	50.71%	Bachillerato Tecnológico
IPN	16	44,902	56.42%	Bachillerato Tecnológico
UNAM	14	185,659	72.29%	Bachillerato General
SE - Telebachillerato Com.	40	278	1.9%	Bachillerato General
SE - Prep. Oficiales	175	28,137	40.66%	Bachillerato General
Ténica	101	8,684	20.81%	Profesional Técnico
SE - Col. Bachilleres	20	2,572	10.08%	Bachillerato General
SE - C. Est. Científicos y Tecn.	57	7,973	18.11%	Bachillerato Tecnológico
SE - C. B. Tecnológico	110	11,088	20.3%	Bachillerato Tecnológico
UAEM	1	1,259	2.05%	Bachillerato General
Tec. del Mar	1	34	0.66%	Bachillerato Tecnológico
Total	700	325,403		

Notes: This table shows the number of educational options per sub-system and how many applicants listed each subsystem as their first option. The column *in preferences* shows the percentage of the applicants' rank-ordered lists with at least one educational option of the subsystems. The last column specifies what type of education each subsystem offers. "Bachillerato General" offers material on various topics, such as mathematics and social sciences, which complements the general knowledge acquired during secondary and primary school. "Bachillerato tecnológico" offers material on various topics, similar to "Bachillerato General", but it also includes technological courses and the opportunity to get a technician degree. "Profesional técnica" offers education of a specialized character in many careers or professions at the upper secondary level while it also includes general high school education into its curriculum. For more information on the type of education, see the *COMIPEMS*' webpage.

The admission process for the MAMC high school choice mechanism is as follows. First, the students must submit a rank-ordered list of up to 20 educational options available in that year's process.[5] Secondly, the students present an exam of 128 questions that cover different topics, such as math, chemistry, Spanish, and history, among others. There are two versions of the admission exam. One version is elaborated by *CENEVAL* (*Centro Nacional de Evaluación para la Educación*

[5]The list of available options is made public by COMIPEMS through an instruction manual published after the announcement of the contest. In this instruction manual, each sub-system specifies the conditions to be accepted (for example, minimum middle school GPA) and the type of education they offer. Also, the instruction manual groups the available options by municipality.

Superior) and the other by *UNAM*. Only the students that choose an *UNAM*'s option as their first will take *UNAM* version. Thirdly, before the matching algorithm occurs, a computer program generates a database that contains the students' exam scores, names, folios, rank-ordered lists, middle school GPAs, and the number of available seats and particular requirements of each educational option. Lastly, *COMIPEMS* implements the Serial Dictatorship algorithm using the exams' scores as a priority (Pariguana and Ortega-Hesles, 2022):

- Assign the student with the highest priority (i.e., the highest score in the exam) to her most preferred educational option.
- After removing from the list the assigned students in the previous step, continue with the student with the highest priority and assign her to her most preferred *available* educational option. *Availability* is determined by the number of seats in the educational option that have not yet been assigned. The student is unmatched if all her ranked educational options do not have an available seat to offer.
- Repeat the previous step until all the students have been processed.

The only priority to rank the students is the exam score, and multiple students could have the same exam score. During the implementation of the previous process, several students often have the same priority (exam score), and, given their preferences and available educational options, they demand more seats than those available in a given educational option. In this case, a representative from the sub-system to which such educational option belongs must decide if all the tied students are accepted or rejected. This means the number of admitted students at each school could be less, equal, or more than the initial number of available seats. This tiebreak decision occurs during an event organized by COMIPEMS since the algorithm is path-dependent, so the tiebreak decision affects subsequent assignments. Finally, after the algorithm and tiebreaks are implemented, some students might be left unassigned, and some educational options might have available seats. In the final stage, unassigned students[6] can apply to options with available seats. The students with higher exam scores are prioritized in this last phase.

In a similar fashion to Pariguana and Ortega-Hesles (2022), we argue that the reported rank-ordered lists reveal the students' truthful preferences. First, the algorithm implemented by *COMIPEMS*

[6]These students are called *Applicant with Right to Choose another Option* (CDO, for its initials in Spanish)

is not strategy-proof because there is a limit on how many schools each student can list in their rank-ordered lists. However, only 6.5% of all students in the 2017 contest reported 20 schools in the rank-ordered lists. Thus, the restriction is not active for the majority of the students. Secondly, because the students do not know their priorities indexes (exam scores) before filling out the rank-ordered lists, they do not know their exact probabilities of admission at each of the ranked schools, which incentivizes truthful revelation of preferences (Pariguana and Ortega-Hesles, 2022).

2.2 The 2017 error

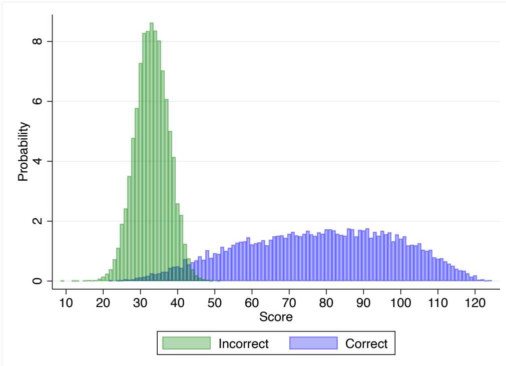
In 2017, around 14,046 students (4.3% of total applicants) obtained an incorrect exam score due to an error while grading the admission exams (Rebolledo, 2017). In particular, the error occurred to some students who took the *UNAM*'s exam version (57% of them). The error was noticed after the publication of the 2017 assignments. Consequently, the authorities decided that the situation had to be revised for those students who received an incorrect exam score. The authorities reported that the error was caused by a gap between some templates of questions and answers (Aristegui Noticias, 2017).

To reassign the students with an incorrect exam score, the authorities used the options' exam score cut-offs[7] of the original assignment (i.e., the assignment generated using the incorrect exam scores) to assign the incorrectly-graded students. For example, if the corrected score of an affected student i is x_i, she would be assigned to the option highest in her rank with an uncorrected-scores cut-off value lower or equal to x_i.

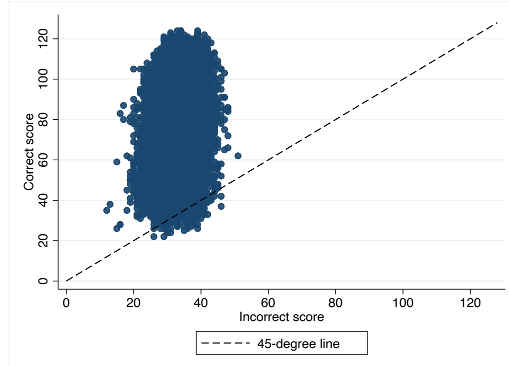
In *Figure 1*, we show the distributions of the exam scores and a scatter plot of the first and second scores of the students with an incorrect initial score. Most students with an incorrect exam score received a lower score before *COMIPEMS* corrected the error. Also, as shown in *Figure 2*, the group of students with incorrect initial exam scores had higher average scores than the rest after the correction. In *Table 2*, we show that, on average, the affected and no affected students are quite different, which indicates that the error was not random. For example, the average second score and middle school GPA for the students with an initial incorrect score is approximately 13% and 2% points higher, respectively, than for the rest of the students.

[7]This is equal to the lowest exam score among the students assigned to such option.

Figure 1: Exam scores of students with incorrect first scores



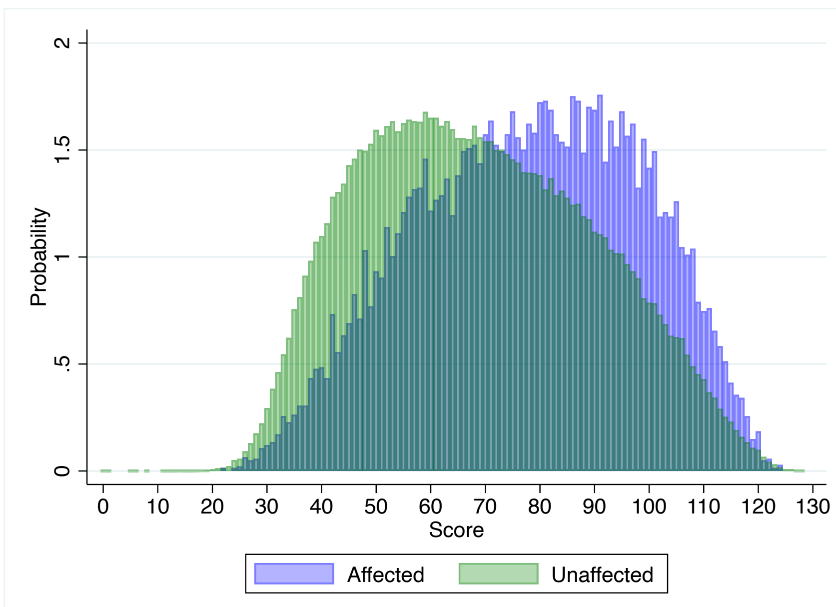
(a) Scores' histograms



(b) Comparison between first and second scores

Note: The error affected the first score of around 14,000 students during grading. Panel (a) shows histograms of the incorrect exam scores (green) and the correct exam scores (blue). Panel (b) shows a scatter plot between the incorrect and correct scores and a 45° line. Notice that, for most students, the incorrect score was lower than the correct score.

Figure 2: Distribution of correct test scores for students affected and unaffected by the grading error



Note: This graph compares the distribution of correct exam scores across affected and unaffected students. The affected students (N = 13,712) are those students who received an incorrect initial exam score, while the unaffected students (N = 290,651) are those who received a correct initial exam score.

The error caused several individuals to get different match outcomes compared to the match outcomes that would have resulted had the error never occurred. To show the welfare impact of the error, in section 4, we simulate the match outcomes assuming that the error never occurred. With

Table 2: Characteristics of affected and unaffected students

Affected Variable	(1) NO Mean/SE	(2) YES Mean/SE	T-test P-value (1)-(2)
First Score	69.213 (0.039)	33.100 (0.038)	0.000***
Second Score	69.213 (0.039)	78.238 (0.175)	0.000***
Middle School GPA	8.101 (0.002)	8.263 (0.007)	0.000***
Male	0.497 (0.001)	0.453 (0.004)	0.000***
Indigenous	0.058 (0.000)	0.053 (0.002)	0.005***
Private middle school	0.077 (0.000)	0.107 (0.003)	0.000***
Morning Shift (middle school)	0.734 (0.001)	0.764 (0.004)	0.000***
Middle school scholarship	0.102 (0.001)	0.110 (0.003)	0.001***
M. Middle school or less	0.543 (0.001)	0.460 (0.004)	0.000***
M. High school or technical	0.330 (0.001)	0.372 (0.004)	0.000***
M. Bachelor's degree or more	0.126 (0.001)	0.167 (0.003)	0.000***
F. Middle school or less	0.539 (0.001)	0.464 (0.004)	0.000***
F. High school or technical	0.317 (0.001)	0.343 (0.004)	0.000***
F. Bachelor's degree or more	0.144 (0.001)	0.194 (0.003)	0.000***
N	290,651	13,712	

Notes: The table compares some of the characteristics of the affected and unaffected students. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. F. stands for father and M. for mother.

those simulations, we are able to see the impact of the mistake in the match outcomes of the 2017 contest and compute welfare differences between the observed match outcomes and the simulated ones.

3 Administrative Data

We use individual-level administrative data from the admission process in 2017. That year, 325,403 students registered, and 700 educational options were available. Our information includes the students' admission exam scores (with and without error), rank-ordered lists, assignments (before and after the error recognition), and middle school GPA. We also observe socioeconomic information reported by the students when they submit their rank-ordered lists. This data includes parents' education levels, information on house assets such as the number of televisions and computers, and home addresses. Table 3 shows summary statistics of socioeconomic variables and admission exam scores.

Table 3: Summary Statistics

	mean	sd	p5	p50	p95
First score	67.59	22.01	34	66	105
Second score	69.62	21.21	37	68	106
Length of rank-ordered list.	10.22	4.29	4	10	20
Middle school GPA	8.11	.87	6.8	8.1	9.6
Man	.5	.5	0	0	1
Indigenous	.06	.23	0	0	1
Private middle school	.08	.27	0	0	1
Morning shift (middle school)	.74	.44	0	1	1
Middle school scholarship	.1	.3	0	0	1
Mother's education:					
Middle school or less	.54	.5	0	1	1
Highschool or technical	.33	.47	0	0	1
Bachelor's degree or more	.13	.33	0	0	1
Father's education:					
Middle school or less	.54	.5	0	1	1
Highschool or technical	.32	.47	0	0	1
Bachelor's degree or more	.15	.35	0	0	1

Notes: The total number of observations is 304,363 (only includes students eligible to participate in the assignment's algorithm). Man, indigenous, education level, and morning shift are indicator variables.

We also have information on the number of seats available at each school, which is required to run the assignment algorithm. With the rules defined by *COMIPEMS*, we can replicate the assignments that *COMIPEMS* made and generate new ones by changing the students' exam scores.

To verify that assignments were made according to the assignment algorithm, we attempt to replicate assignments using the initial 2017 test scores. We are able to replicate 99.99% of the original match outcomes of 2017[8]. Three types of participants enrolled in the 2017 admission process: i) those who were assigned to a school during the algorithm (matched), ii) those who were not assigned to a school during the algorithm (unmatched), and iii) those students who committed an infraction while taking the exam, did not take the exam, or did not present their middle school certificate on time (ineligible).

To clarify things, from now on, we give the following names to the match outcomes that we observe in our data:

- Original: the first assignments made by*COMIPEMS*, using the incorrect exam scores for the affected students.
- Reassignment: the second assignments made by *COMIPEMS* after correcting the test scores. The last statement is data-driven because *COMIPEMS* did not specify in its declarations how exactly they corrected the error.

Table 4 shows some statistics of the two match outcomes observed in our data. Around 253,000 (77.87%) of eligible students were assigned to a school (matched) during the original assignment. After the correction was made (reassignment) and the students that were unmatched applied to the educational options with available seats (CDO), the number of matched students increased to 268,498, which is an increase equivalent to approximately 5% of the total number of applicants (325,403).

[8]We believe that we are not able to replicate the 100% of the match outcomes due to an error in the reported middle school GPA of two students. We know about the possible error in their middle school GPA because *COMIPEMS* changed their assignments in the re-assignment phase, which is consistent with what we obtained in our replication of the original match outcomes.

Table 4: Match Outcomes

Type		Original N	Original %	Reassignment + CDO N	Reassignment + CDO %
Matched		253,383	77.87	268,498	82.51
Unmatched		50,980	15.67	36,277	11.15
Ineligible	Infraction	5	0.00	5	0.0
	No exam	11,262	3.46	11,262	3.46
	No middle school	9,773	3.00	9,361	3.00
Total		325,403	100	325,403	100

Note: The match outcomes are from the original assignment in 2017. The difference between the "original" and "reassignment" columns in the "No middle school" column occurs because 412 students who did not have their middle school certificate at the start of the process were assigned to an educational option during the CDO phase. Ineligible applicants were dismissed because they did not present the admission exam, did not turn in their middle school certificate, or committed an infraction while taking the admission exam.

To complement the individual-level data, we use the addresses reported by the students to generate student coordinates (for more detail, see appendix A) and obtained the schools' coordinates from the official website of *COMIPEMS* (for more detail see appendix B). Finally, we generated a matrix containing the distances from each student's house to the 700 schools participating in the 2017 admission process using the students' and schools' coordinates.

4 Simulations

To understand the impact of the exam grading error on welfare, we need to understand how it affected the outcomes of students' matches. However, we do not observe the match outcomes the students should have received without the grading error. This scenario never happened because the authorities published the match outcomes before knowing about the error, and once they became aware of the error, they only revised the match outcomes of the students with incorrect exam scores. Therefore, we need to simulate the match outcomes that would have occurred without the grading error. To conduct this simulation, we use *COMIPEMS*'s decision rules explained in section 2 with the full set of correct exam scores. The resulting match outcomes of the simulation represent the assignments that the participants would have received if the error had never occurred.

A part of the assignment process is not automatic or solely determined by the algorithm. The assignment uses the admission exam scores as a priority index (a higher exam score implies higher priority). Still, several students could have the same exam score and, therefore, the same priority. This characteristic of the assignment mechanism implemented by *COMIPEMS* requires the intervention of sub-system managers: the manager or the representative of each educational option must decide to accept or reject all tied students depending on the number of available seats. For example, if student i and k have the same exam score and applied to school j, which has only one seat available, the management/representative of school j must decide whether to accept or reject both participants. This tie-breaking decision is not algorithm-based.

The managers' intervention during the implementation of the assignment mechanism complicates the simulations as we do not know what decision they would have made. We can assume that each school's decision on the original assignment - to accept or reject all tied students - would be the correct assumption for the simulation, but this may be incorrect. As we change the exam scores that determine the priority index of the participants, the number of tied students could change for some schools. The probability of accepting or rejecting all the tied students can be represented as a function of the "excess" demand of tied students for the school. As the number of tied students increases, it is intuitive that the probability of accepting all tied students decreases. For example, accepting a higher-than-anticipated number of students could generate budget or logistical complications. On the other hand, if the excess demand tends to zero (the number of tied students is equal to the number of available seats at the school), the probability of accepting all the tied students tends to one.

Therefore, we generate three simulations under different assumptions on how the schools deal with ties and excess demand:

- Unfavorable tie decision: Use each school's initial offer of available seats in 2017 as the school's capacity and the corrected exam scores as inputs. Run the assignment algorithm. If some students with the same admission exam score compete for the last available seats at a school, that school rejects all the tied students. With this simulation, we force the number of accepted students for all the schools to be less or equal to the initial offer of seats.

- Favorable tie decision: Use each school's initial offer of available seats in 2017 as the school's capacity and the corrected exam scores as inputs. Run the assignment algorithm. If some students with the same admission exam score compete for the last available seats at a school, that school accepts all the tied students. With this simulation, we force the number of admitted students for all the schools to be greater or equal to the initial offer of seats.

- Intermediate tie decision: Calculate the assignments from the favorable and unfavorable tie decision scenarios. From the date, observe the number of students admitted by each school before and after the error correction. If school j accepted more students after the correction than before, choose the "favorable tie decision" scenario for that school. If school j accepted fewer students after the correction than before, choose the "unfavorable tie decision" scenario for that school.

Table 5 shows some statistics of the simulated and observed match outcomes. The number of matched students increases as we are more flexible with the schools' tie decisions. Also, note that the mean ranking of the assigned option is higher in all the simulated scenarios than in the reassignment.

Table 5: Match outcomes

	Simulated			Observed	
	Unfavorable	Intermediate	Favorable	Original	Reassignment
Num. of matched students	242,706	254,933	258,937	253,383	261,683
Num. of unmatched students	61,657	49,430	45,426	50,980	42,680
Mean num. of option assigned	4.84	4.61	4.52	4.5	4.48
% of students assigned first option	18.79%	21.39%	22.25%	22.36%	22.91%

Notes: The table shows the number of students assigned under different scenarios. The simulated match outcomes depend on our assumptions about the schools' tie decisions. The row percentages are for the total number of eligible students (304,361).

After the error was detected, *UNAM*, the entity deemed responsible for the error, stated that they fixed the mistake and assigned all the students with a wrong exam score to the corresponding school based on their correct exam scores and rank-ordered lists (see *UNAM*'s press release). However, they did not specify precisely how they did the correction. After analyzing the data, we concluded that *COMIPEMS*'s approach to solving the mistake was to use the original assignment's exam score cut-offs to match the affected students using their correct exam score.

There are two ways by which the students could get a different match outcome due to the error:

- Directly affected: students with an incorrect exam score who received a reassignment different from the simulated assignment without error (under the unfavorable, favorable, or intermediate assumptions).

- Indirectly affected: students with a correct exam score who received an assignment different from the simulated assignment without error (under the unfavorable, favorable, or intermediate assumptions).

Table 6 shows the number of indirectly and directly affected students under each assumption. The error caused around 14,000 incorrect exam scores. However, because of spillover effects, under the most restrictive assumption, the number of students that got a different match outcome because of the error is almost 80,000 students. Even using the most flexible assumption, the observed and simulated match outcomes differ for approximately 13,000 students. The existence of directly and indirectly affected students implies that the error has welfare implications, which we will compute in the following sections.

Table 6: Affected students

Simulation	Directly	Indirectly	Total affected	% of total
Unfavorable	3,992	74,767	78,759	25.88%
Intermediate	1,681	31,148	32,829	10.79%
Favorable	588	12,349	12,937	4.25%

Notes: This table shows the number of affected students under the different simulations. An affected student is a participant that received a different assignment compared to the "universe" without error (simulations). Furthermore, a directly affected students are those that were affected because they received an incorrect score, while indirectly affected students are those that received a different match outcome due to spill-over effects. The percentage column is with respect of the total of eligible students (304,361).

The number of directly affected students in the second column of table 6 may be larger or smaller than the number of students with incorrect exam scores. This occurs for several reasons. First, the directly affected students with exam scores greater than the unfavorable simulation's exam score cut-offs (the biggest cut-off of the observed and simulated scenarios due to its unfavorable tie decisions) will always get the same assignment under the observed reassignment and the simulated scenarios. Secondly, the directly affected students with exam scores similar to the cut-offs of the favorable simulation could lose their seats in higher-ranking schools as we are more unfavorable with the schools' tie decisions. For example, if a student has an exam score equal to the cut-off of the school in which she was assigned and that school has excess demand (i.e., the number of tied students is higher than the number of available seats), and we change that school's tie decision to reject all tie students, she will not be assigned to that school. The latter is a clear example on why as we are more unfavorable with the schools' tie decisions the number of directly affected students increases. Lastly, the first two facts indicate that the majority of the directly affected students are well above the schools exam score cut-offs and, thus, are not affected by the error when we compare their reassignments with the "universe without error" simulated assignments.

In table 7, we analyze the consequences if the error had not been corrected. By looking at the column "Directly", we can notice that the error had strong consequences for directly affected students, i.e. the error affected the admission exam scores such that the majority of the directly affected students received an incorrect assignment. On the other hand, the indirectly affected students benefited from the error.

Table 7: Affected students

Simulation	Directly	Indirectly	Total affected	% of total
Unfavorable	10,101	74,768	84,869	27.88%
Intermediate	10,649	31,149	47,798	13.73%
Favorable	10,885	12,351	23,236	7.63%

Notes: This table shows the number of affected students under the different simulations. An affected student is a participant that received a different original assignment compared to the "universe" without error (simulations). Furthermore, directly affected students are those that were affected because they received an incorrect score, while indirectly affected students are those that received a different match outcome due to spill-over effects. The percentage column is with respect of the total of eligible students (304,361).

5 Preferences

The primary purpose of this paper is to estimate welfare changes following school assignments. For such purpose, we follow Abdulkadiroğlu et al. (2020)) and use the rank-ordered logit model proposed by Beggs et al. (1981). Several studies have used this model to estimate students' preferences over a set of schools. For example, Laverde (2022) uses a rank-ordered logit to recover parental preferences for schools and "estimate how much of the cross-gap in school achievement can be attributed to the location of students." On the other hand, Pariguana and Ortega-Hesles (2022)) use the rank-ordered logit to approximate the effects on ex-ante welfare of changes in decision rules in the assignment algorithm of Mexico City's high school system.

Let U_{ij} denote student i's utility from enrolling in school j, and let $\mathcal{J} = \{1, ..., J\}$ be the set of schools that participate in the centralized admission process. Then, U_{ij} may be interpreted as the indirect utility associated with student i's attending school j.

In our data, the school ranked first in student i's rank-ordered list is her favorite school, the second one is the second favorite school, and so on. Then, the school ranked first on a student's list is

$$R_{i1} = \underset{j \in \mathcal{J}}{\arg\max}\ U_{ij}.$$

This indicates that the utility that student i obtains from assisting to her favorite school in his rank-ordered list gives her the highest utility among all the alternatives. The second most preferred option must be such that the utility of that school is higher than all the available alternatives excluding the first ranked option (and so on); thus, ranks $2, ..., \ell_i$, where ℓ_i is the length of the list submitted by student i, should satisfy:

$$R_{ik} = \underset{j \in \mathcal{J} \setminus \{R_{im} : m < k\}}{\arg\max} U_{ij}.$$

Abdulkadiroğlu et al. (2020) and Pariguana and Ortega-Hesles (2022), among others, summarize these preferences by fitting a random utility model with parameters that vary according to observed student characteristics. Similarly, we model student i's utility from enrolling in school j as

$$U_{ij} = \delta_{c(X_i),j} - \tau_{c(X_i)} D_{ij} + \eta_{ij}, \tag{1}$$

where the function $c(X_i)$ assigns students to covariate cells based on the variables in the vector X_i, D_{ij} is the distance from i's house to school j, and η_{ij} represents the unobservable component of the utility, which is modeled as an independent extreme value type I distribution conditional on X_i. Equation 1 is a rank-ordered multinomial logit model. The parameter τ represents the importance of proximity for the students, and the parameter δ summarizes the location-independent attractiveness of school j in each covariate cell (Laverde, 2022).

The strategy of dividing the applicants into several groups defined by covariates follows Abdulkadiroğlu et al. (2020), Laverde (2022), and Hastings et al. (2017). In addition to the flexible preference heterogeneity, this strategy allows us to generate tractable versions of the rank-ordered logit because the lower number of parameters reduces the computational power required to estimate them.

The logit model implies that the conditional likelihood of the rank list $R_i = (R_{i1}, ..., R_{i,\ell(i)})$ is

$$\mathcal{L}\left(R_i \mid X_i, D_i\right) = \prod_{k=1}^{\ell(i)} \frac{\exp\left(\delta_{c(X_i)R_{ik}} - \tau_{c(X_i)} D_{i,R_{ik}}\right)}{\sum_{j \in \mathcal{T}\{R_{in}: m<k\}} \exp\left(\delta_{c(X_i)j} - \tau_{c(X_i)} D_{ij}\right)}, \tag{2}$$

where $D_i = (D_{i1}, ..., D_{iJ})$. To allow flexible heterogeneity in tastes, we estimate preference models separately for 32 covariate cells defined by the intersection of the dichotomous variables sex, mother with low education, father with low education, morning shift in middle school, and no automobile possession. With these 32 groups of students, we estimate the fixed effects $\delta_{c(X_i),j}$ for 570 schools out of a total of 700 schools and for all the covariate cells. We define the remaining 130 schools as the outside option and normalize the utility of enrolling in them to zero.[9] The estimated parameters $\hat{\tau}_{c(X_i)}$ are shown in appendix C.

6 Welfare

6.1 Expected Utility

Throughout the paper, we omit uncertainty and describe the following expected utilities as realized:

$$E[U_{ij}|R_i, V_i],$$

where $V_i = \{V_{i1}, ..., V_{il(i)}\}$ is the deterministic part of the utility.

Following Abdulkadiroğlu et al. (2020), the expected utility of the highest-ranked alternative is:

$$E[U_{i1}|R_i, V_i] = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{u_1} \int_{-\infty}^{u_2} \cdots \int_{-\infty}^{u_{J-1}} \left[u_1 \prod_{j=1}^{J} f\left(u_j \mid V_{ij}\right)\right] du_J \ldots du_2 du_1}{\prod_{j=1}^{J-1} P\left(V_{ij} \mid V_{ij} \ldots . V_{iJ}\right)},$$

where $f(u|V) = \exp(V - u - \exp(V - u))$ is the density function of a Gumbel random variable with location parameter V. The latter simplifies to

$$E[U_{i1}|R_i, V_i] = \frac{\prod_{j=1}^{J} P\left(V_{ij} \mid V_{ij} \ldots V_{iJ}\right) \times \mathcal{I}\left(V_{i1} \ldots V_{iJ}\right)}{\prod_{j=1}^{J-1} P\left(V_{ij} \mid V_{ij} \ldots V_{iJ}\right)} \tag{3}$$

where $\mathcal{I}(V_{i1}, ...V_{iJ}) = \mu_\eta + \log(\sum_{v \in V_i} \exp(v))$.

[9] These 130 schools are those not listed by at least one student in all the 32 covariate cells.

Equation 3 is a direct result of the irrelevance of independent alternatives, and it is similar to the expected utility of a multinomial logit model (Abdulkadiroğlu et al., 2020; Beggs et al., 1981).

Now, for options $j > 1$, define the following functions:

$$G_{i0}(u) = 1,$$
$$f\left(x \mid V_{ik}\right) G_{i(k-1)}(x) dx, k = 1 \ldots J.$$

It can be shown that

$$G_{ik}(u) = \int_{u}^{\infty} f\left(x \mid V_{ik}\right) G_{i(k-1)}(x) dx, k = 1 \ldots J.$$
$$G_{ik}(u) = \sum_{j=1}^{k} B_{ik}^{j}\left[1 - F\left(u \mid \mathcal{I}\left(V_j \ldots V_k\right) - \mu_\eta\right)\right]$$

where $F(u \mid V) = \exp(-\exp(V - u))$ is the Gumbel CDF with location V, and the coefficients B_{ik}^{j} are:

$$B_{i1}^{1} = 1,$$
$$B_{ik}^{j} = -B_{i(k-1)}^{j} \times P\left(V_{ik} \mid V_{ij} \ldots V_{ik}\right), k > 1, j \neq k,$$
$$B_{ik}^{k} = \sum_{j=1}^{k-1} B_{i(k-1)}^{j}, k > 1.$$

Then for $j > 1$, we have

$$\begin{aligned} E[U_{ij}|R_i, V_i] &= \frac{\int_{-\infty}^{\infty} \int_{u_j}^{\infty} \int_{u_{j-1}}^{\infty} \cdots \int_{u_2}^{\infty} \int_{-\infty}^{u_j} \int_{-\infty}^{u_{j+1}} \cdots \int_{-\infty}^{u_{J-1}} \left[u_j \prod_{k=1}^{J} f\left(u_k \mid V_{ik}\right)\right] du_J \ldots du_{j+1} du_1 \ldots du_j}{\prod_{k=1}^{J-1} P\left(V_{ik} \mid V_{ik} \ldots V_{iJ}\right)} \\ &= \frac{\int_{-\infty}^{\infty} u_j f\left(u_j \mid \mathcal{I}\left(V_{ij} \ldots V_{iJ}\right) - \mu_\eta\right) G_{i(j-1)}\left(u_j\right) du_j}{\prod_{k=1}^{j-1} P\left(V_{ik} \mid V_{ik} \ldots V_{iJ}\right)} \\ &= \frac{\sum_{m=1}^{j-1} B_{i(j-1)}^{m}\left[\mathcal{I}\left(V_{ij} \ldots V_{iJ}\right) - P\left(V_{ij} \ldots V_{iJ} \mid V_{im} \ldots V_{iJ}\right) \mathcal{I}\left(V_{im} \ldots V_{iJ}\right)\right]}{\prod_{k=1}^{j-1} P\left(V_{ik} \mid V_{ik} \ldots V_{iJ}\right)} \end{aligned}$$

The last equation occurs because we are conditioning on R_i (the submitted list by the student), which restricts the space where η_{ij} can be (Abdulkadiroğlu et al., 2017; Pariguana and Ortega-Hesles, 2022).[10]

6.2 Welfare changes

To compute the changes in welfare, we compare the utility of the students in different assignments:

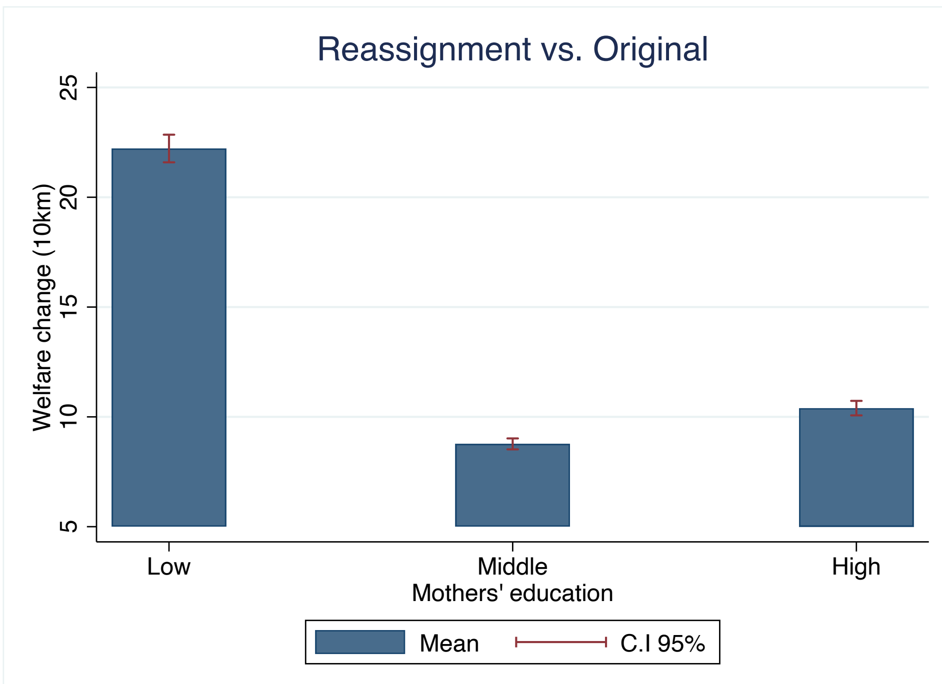
[10]The expected utility of three students can not be computed due to numerical and precision issues.

$$\Delta U_i(assignment_1, assignment_2) = \frac{E[U_{ij}|assignment_1] - E[U_{ij}|assignment_2]}{-\beta},$$

where β is the estimated coefficient for distance in the rank-ordered logit model.[11] Also, we assume that the student's utility equals zero if the student is unmatched. The assumption is not a problem because more than 80% of students are matched in the observed and simulated scenarios.

First, in figure 3, we show the welfare implications of *COMIPEMS* correction by comparing the reassignment and the original assignment for the students with incorrect exam scores that received an incorrect assignment due to the error. We show the mean with 95% confidence intervals.

Figure 3: Welfare change's of the correction



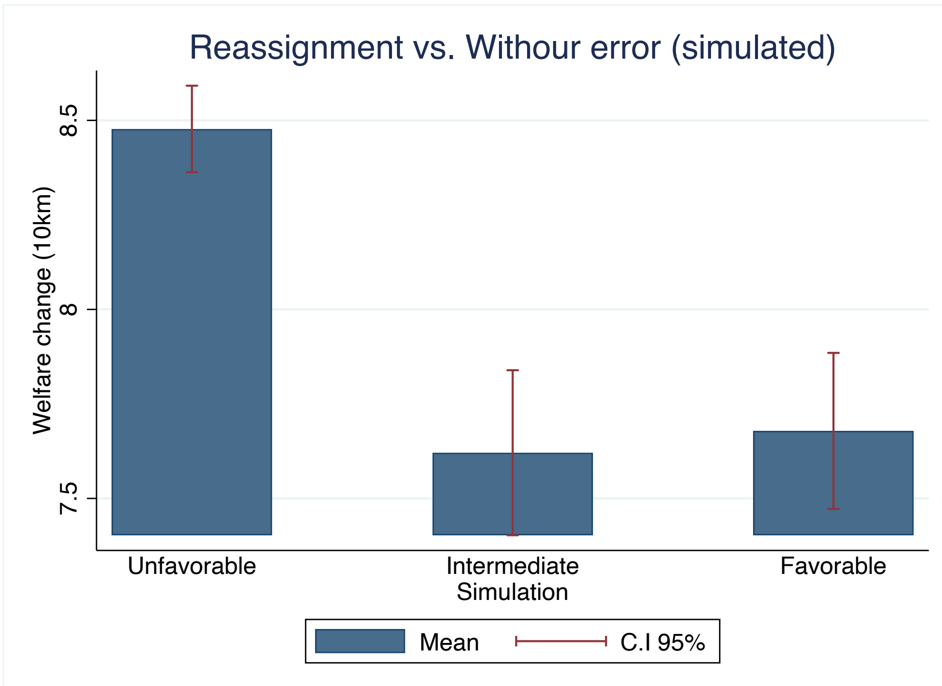
Note: This figure shows the welfare change for the students that received a new assignment after *COMIPEMS* noticed the error.

Secondly, in figure 4 you can see the results of comparing the welfare obtained from the reassignment and the simulations. The average mean welfare change for the unfavorable, intermediate, and favorable assumptions are 8.47, 7.62, and 7.67, respectively. The latter is equivalent to decreasing the distance from the students' houses to schools by 70km approximately. Figure 5 shows the welfare

[11] Dividing the utility by the distance coefficient allows us to interpret the change in welfare in 10km units. Thus, an increase of 1 in utility can be interpreted as a reduction in the distance students travel to the schools by 10km.

mean by assumption if the error had not been corrected. Since the majority of the students were affected directly, as we are more unfavorable in the simulations, the mean welfare change increases. This indicates that several applicants are worse-off under the error assignments than the most favorable scenario.

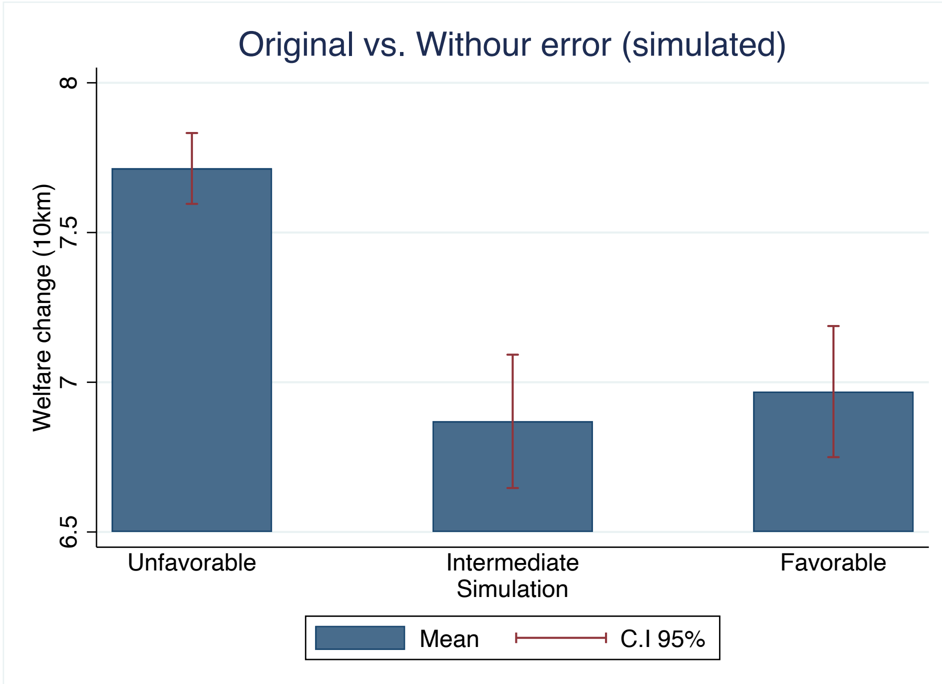
Figure 4: Welfare change's distribution by assumption for all affected students (directly and indirectly)



Note: This figure shows the welfare change by assumption for the affected students (directly and indirectly).

On the other hand, figure 6 shows the welfare distribution by type of affected student under the three simulations. In general, across the three simulations, the median are approximately the same between the directly and indirectly affected students. The affected students distribution is similar to the indirectly affected's one because the number of indirectly affected students represents the higher proportion of all the affected students (see table 6).

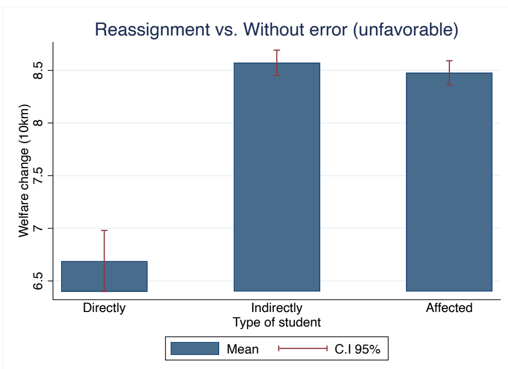
Figure 5: Welfare change's distribution by assumption for directly affected students



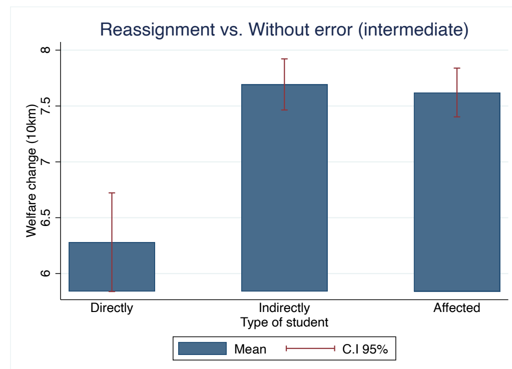
Note: This figure shows the welfare change for the directly affected students by assumption.

Furthermore, to analyze the impact of the error on different socioeconomic groups, we group the students by their mother's education (low, mid, and high). On appendix E you can find box plots of the welfare change by groups under the different assumptions (min, int, and max). Also, to show the impact of the error, we use only the affected students as sample to generate the box plots. Notice that, for the three simulations, a higher mothers' education is correlated negatively with the median welfare change. Thus, the low socioeconomic groups were the most benefited by the error.

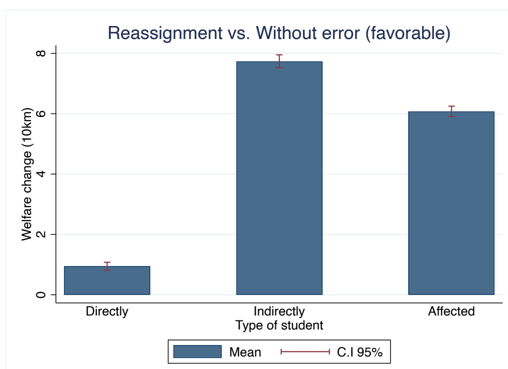
Figure 6: Welfare change's distribution by type of affected student



(a)

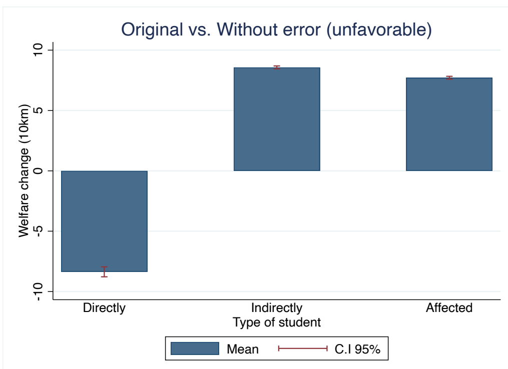


(b)

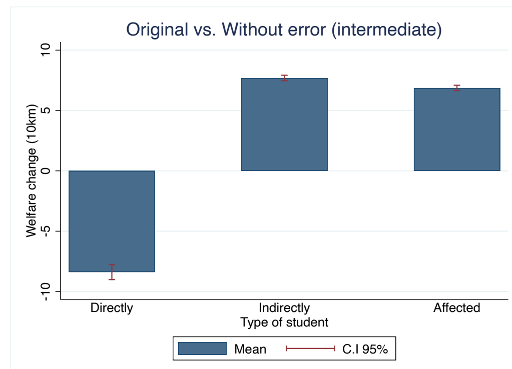


(c)

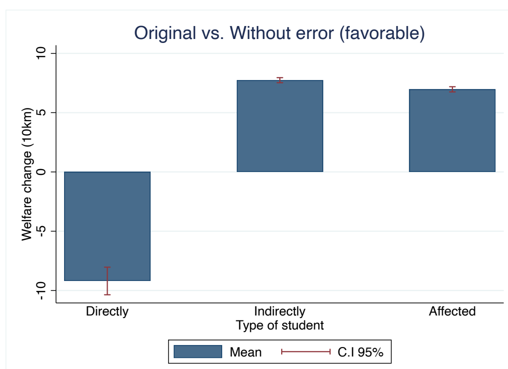
Figure 7: Welfare change's distribution by type of affected student



(a)



(b)



(c)

7 Conclusion

The main contribution of this paper is to show the consequences of a human error during centralized school assignment mechanisms. In particular, we analyze the 2017 error in the centralized contest of public high schools in the Metropolitan Area of Mexico City. The error affected the admission exam scores (which are the students' priorities in the assignment algorithm) of 14,000 students. Using the administrative data of that year, we were able to simulate scenarios in which the error had never occurred and compared them with the final assignments observed. We generate a low and high boundary for the total number of assignments that changed due to the error compared to those of an "universe" without error. Between 12,349 and 74,767 students, depending on the breaking ties assumptions, received a different assignment. This shows that an error can affect more students indirectly through spillover effects.

Furthermore, we compute the welfare implications of the error. The median welfare change for the affected students (those that received an assignment different to that that should had happened without error) is equivalent to decreasing the students' distance to the schools by approximately 70 kilometers.

However, due to limitations on the administrative data, we do not estimate the impact of the error on variables such as school dropout or lower average GPA. This is relevant because, as is shown by Dustan et al. (2017), been admitted to a "elite school" increases the probability of high school dropout by 9.4 percentage points, while only increasing modestly the end-of-high-school math test scores for the marginal admittee. Thus, an interesting question for future research could be analyzing the impact of the error on outcome variables such as average GPA or high-school dropout.

References

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

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

Abdulkadiroğlu, A. and T. Sönmez (2003, June). School choice: A mechanism design approach. *American Economic Review 93*(3), 729–747.

Aristegui Noticias (2017). Por error en revisión de exámenes, comipems publicará nueva lista de resultados. https://aristeguinoticias.com/0708/mexico/por-error-en-revision-de-examenes-comipems-publicara-nueva-lista-de-resultados/. Accessed: 2022-09-20.

Beggs, S., S. Cardell, and J. Hausman (1981). Assessing the potential demand for electric cars. *Journal of Econometrics 16*, 1–19.

Dustan, A., A. D. Janvry, and E. Sadoulet (2017). Flourish or fail? the risky reward of elite high school admission in mexico city. *The Journal of Human Resources 52*(3), 756–799.

Elacqua, G., I. Jacas, T. Krussig, C. Mendez, C. Neilson, A. Roman, and S. Soares (2021). Centralized student assignment systems. Technical Report 2184.

Hastings, J., A. Hortaçsu, and C. Syverson (2017). Sales force and competition in financial product markets: The case of mexico's social security privatization. *Econometrica 85*(6), 1723–1761.

Laverde, M. (2022). Distance to school and equal access in school choice systems.

Manjunath, V. and B. Turhan (2016). Two school systems, one district: What to do when a unified admissions process is impossible. *Games and Economic Behavior 95*, 25–40.

Ngo, D. and A. Dustan (2022). Preferences , access , and the stem gender gap in centralized high school assignment.

Pariguana, M. and M. E. Ortega-Hesles (2022). School choice, mismatch, and graduation.

Rebolledo, R. (2017). 10 cosas que debes saber sobre el error de comipems. *El Economista*.

Appendices

A Students' Coordinates

The data from COMIPEMS contains the addresses that all the students registered while filling out the application forms. However, several observations contain missing or, presumably, incorrect information. The applicants' information reported are addresses, municipalities, and zip codes. With these variables, we generated a string variable containing the complete address information.

With the variable generated, we used the API from HERE to get the coordinates of all the applicants. The results from the web API of HERE contain a variable called "type", which, in our case, is a metric of how specific/accurate is the localization of the address. The results are the following:

Table 8: Result Types

Type	Obs.	Percentage
administrativeArea	2	0%
houseNumber	245,533	75%
intersection	1,361	0.4%
locality	25,400	8%
place	3,437	1%
street	34,522	11%
Total	310, 255	95%

Notes: The percentage does not sum up to 100% because some addresses were not located by HERE's API. The coordinates are the centroids.

The result of type "administrativeArea" is when the search returns a very general location (Country, state, city). The "houseNumber" result implies that the search located the address at a house level. The "intersection" result is the coordinates of the intersection between two streets. The "locality" result is the centroid of the address's borough. The "place" result is when the search assigns the address to a specific place, such as a restaurant or a church. The "street" results return the centroid of the street associated with the address. Another metric that the HERE API returns that can be useful to determine the results' reliability is a score between 0 and 1, where 1 implies a perfect localization of the address.

After using the HERE API, we have 20,113 participants with missing or incorrect coordinates. Therefore, We used Google Maps API to extract the missing coordinates and see if the results were better or at least consistent and obtained coordinates for 17,762 participants. For the other 2,351 participants, the Google API was not able to return a result or returned coordinates outside of Mexico (see table 9). The description of each type of result is as follows:

- Approximate: returns only the addresses that are characterized as approximate.
- Geometric Center: returns only geometric centers of a location such as a polyline (for example, a street) or polygon (region).
- Range Interpolated: returns only the addresses that reflect an approximation (usually on the road) interpolated between two precise points (such as intersections). An interpolated range generally indicates that rooftop geocodes are unavailable for a street address.
- Rooftop: returns only the addresses for which Google has location information accurate down to street address precision.

Google Maps API does not return the country of the coordinate extracted; however, using spData geocoordinates from all the countries in the world, we were able to identify the observations that are in Mexico (see figure 8). Therefore, we end up with 323,052 (99.27% of the total sample of 325,403) participants with coordinates in Mexico (using either Here or Google's API).

Table 9: Result Types Google

Type	Obs.	Percentage
No coords	2,351	12%
Approximate	4,865	24%
Geometric Center	7,676	38%
Range Interpolated	1,603	8%
Rooftop	3,618	18%
Total	20,113	100%

Notes: To see a more general overview of the results that Google Maps' API could return visit its web page.

Figure 8: Coordinates obtained with Google Maps API

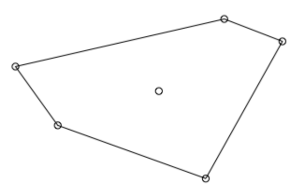
(a) All results

(b) Mexico results

Notes: The figure shows the spatial distribution of the students' addresses obtained using Google Maps API.

We have 2,351 participants that do not have coordinates because Google/HERE could not identify the addresses that the students gave to COMIPEMS. For these observations, We used GeoNames data. The data contains zip codes' coordinates. However, some zip codes can have multiple points. For example, one zip code is the same for multiple suburbs; thus, the data have an entry for each suburb with different coordinates. To avoid more data processing, We generated a convex hull with the different coordinates per zip code and computed the centroid of that convex hull (see figure 9 for an example).

Figure 9: Convex hull and its centroid

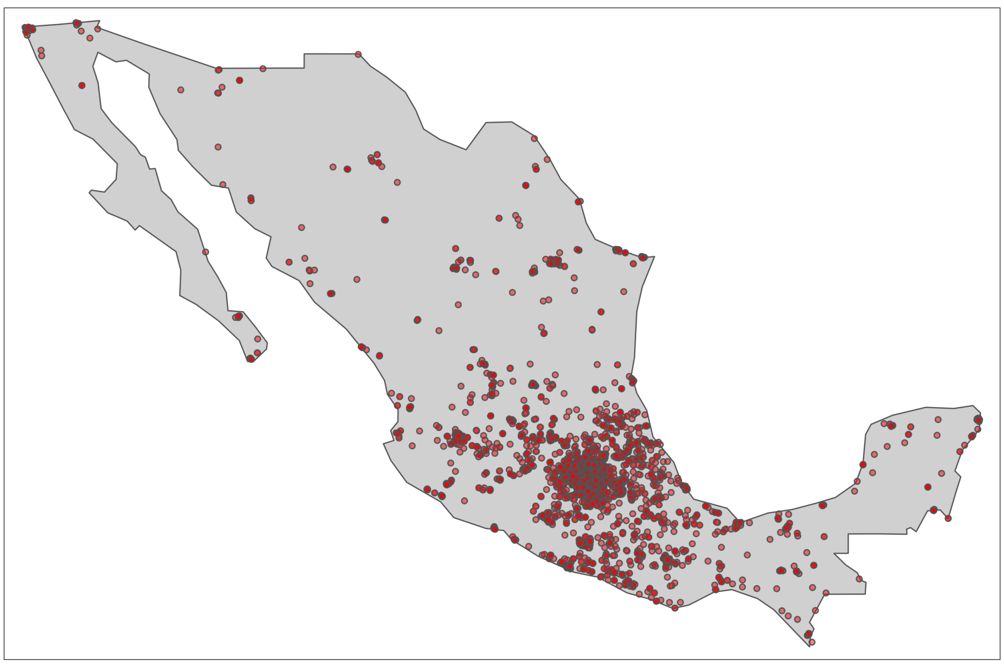


Before merging the data sets from GeoNames and our data, We added a 0 to the start of the zip codes that only have four digits (zip codes should have five digits). One explanation of why some zip codes reported by the participants have less than five digits is that students wrote "4444" when the zip code is "04444". Some participants reported zip codes with only 3 and 2 digits. We assumed that the participants who reported a four-digit zip code did because of the start-with-zero.

Thus, we added to those observations a 0 at the start [12]. Then, we merged both data sets. Of the 2,351, 38 observations that did not have a zip code that existed in the GeoNames data. We used Google Maps API for these observations and extracted the coordinates (the centroid) for the *delegación* that the participants reported. Finally, after using the centroid of the *delegación*, only two observations reported a *delegación* that does not exist. For example, a participant reported as *delegación* "Chapultepec". So, we have only two missing values that we manually searched.

In figure 10 you can see the geographical distribution of the students that participated in COMIPEMS's 2017 contest.

Figure 10: Coordinates of the addresses reported by the 2017 contest's participants (only Mexico)



Notes: The figure shows the spatial distribution of the students' reported addresses in Mexico. All the coordinates shown were obtained from Google Api, HERE API or manually.

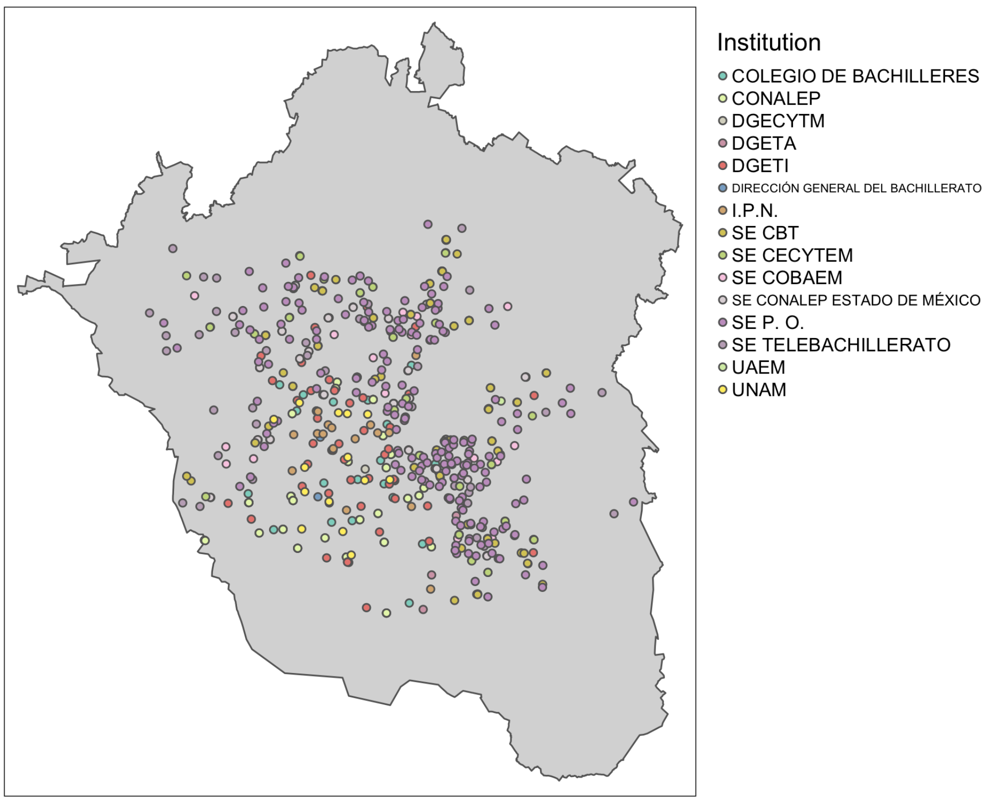
B Schools' Coordinates

The schools' coordinates were extracted from the official website of COMIPEMS using the identifiers contained in our administrative data. Unfortunately, 28 programs could not be located on the website of COMIPEMS since the coordinates correspond to the 2021 list of school options. Nevertheless, we

[12] Some of these observations are mainly at Mexico City, which is consistent since zip codes starting with 0 are from Mexico City

obtained the coordinates of 25 of the 28 programs missing using the campus' name. The number of schools with coordinates available is 697 [13]. See figure 11 to see the schools' geographic distribution.

Figure 11: Schools' geographic distribution in *Zona Metropolitana del Valle de México*



Notes: The figure shows the geographic distribution of the schools that participated in the 2017 *COMIPEMS* contest.

[13] The identifier of the schools with no coordinates are S266000, S268000, and S866000. 4, 3, and 147 students were assigned to each school, respectively. We use the coordinates of the centroid of the postal code reported by these schools.

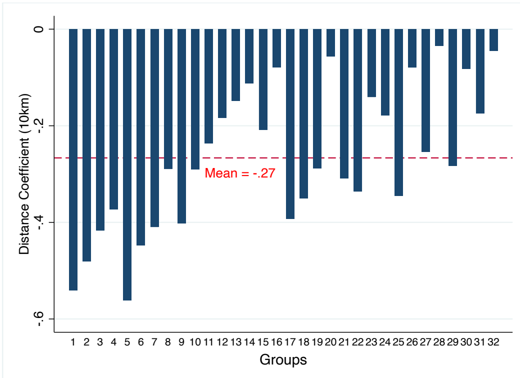
C Rank-ordered logit estimates

Table 10: DISTANCE'S COEFFICIENTS

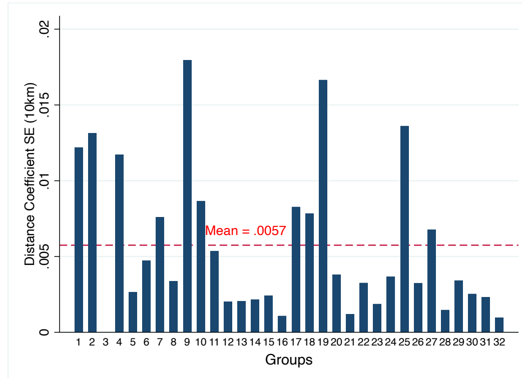
Group	Man	M. middle or less	Morning S.	F. Middle or less	No car	Coef.	SE
1	NO	NO	NO	NO	NO	-.54096	.01221
2	NO	NO	NO	NO	YES	-.481	.01314
3	NO	NO	NO	YES	NO	-.4172	0.02011
4	NO	NO	NO	YES	YES	-.37284	.01172
5	NO	NO	YES	NO	NO	-.56116	.00266
6	NO	NO	YES	NO	YES	-.44767	.00474
7	NO	NO	YES	YES	NO	-.40973	.00759
8	NO	NO	YES	YES	YES	-.2896	.00337
9	NO	YES	NO	NO	NO	-.40219	.01795
10	NO	YES	NO	NO	YES	-.28985	.00867
11	NO	YES	NO	YES	NO	-.2364	.00537
12	NO	YES	NO	YES	YES	-.18337	.00202
13	NO	YES	YES	NO	NO	-.14803	.00205
14	NO	YES	YES	NO	YES	-.11183	.00217
15	NO	YES	YES	YES	NO	-.20828	.00243
16	NO	YES	YES	YES	YES	-.07922	.00109
17	YES	NO	NO	NO	NO	-.39239	.00828
18	YES	NO	NO	NO	YES	-.3504	.00785
19	YES	NO	NO	YES	NO	-.28786	.01665
20	YES	NO	NO	YES	YES	-.05674	.0038
21	YES	NO	YES	NO	NO	-.30857	.00121
22	YES	NO	YES	NO	YES	-.33627	.00326
23	YES	NO	YES	YES	NO	-.14064	.00187
24	YES	NO	YES	YES	YES	-.1788	.00368
25	YES	YES	NO	NO	NO	-.34528	.01362
26	YES	YES	NO	NO	YES	-.0795	.00325
27	YES	YES	NO	YES	NO	-.25377	.00678
28	YES	YES	NO	YES	YES	-.03471	.00148
29	YES	YES	YES	NO	NO	-.28319	.00342
30	YES	YES	YES	NO	YES	-.08186	.00254
31	YES	YES	YES	YES	NO	-.17468	.00232
32	YES	YES	YES	YES	YES	-.04497	.00098

Notes: The table shows the distance's coefficient estimation for the 32 groups generated using five variables. Table for coefficients of 10km. M == Mom, F == Father, Morning == Morning Shift in middle school. Error in the approximation of the group 3's SE.

Figure 12: Coefficients and Standard Errors



(a) Coefficients

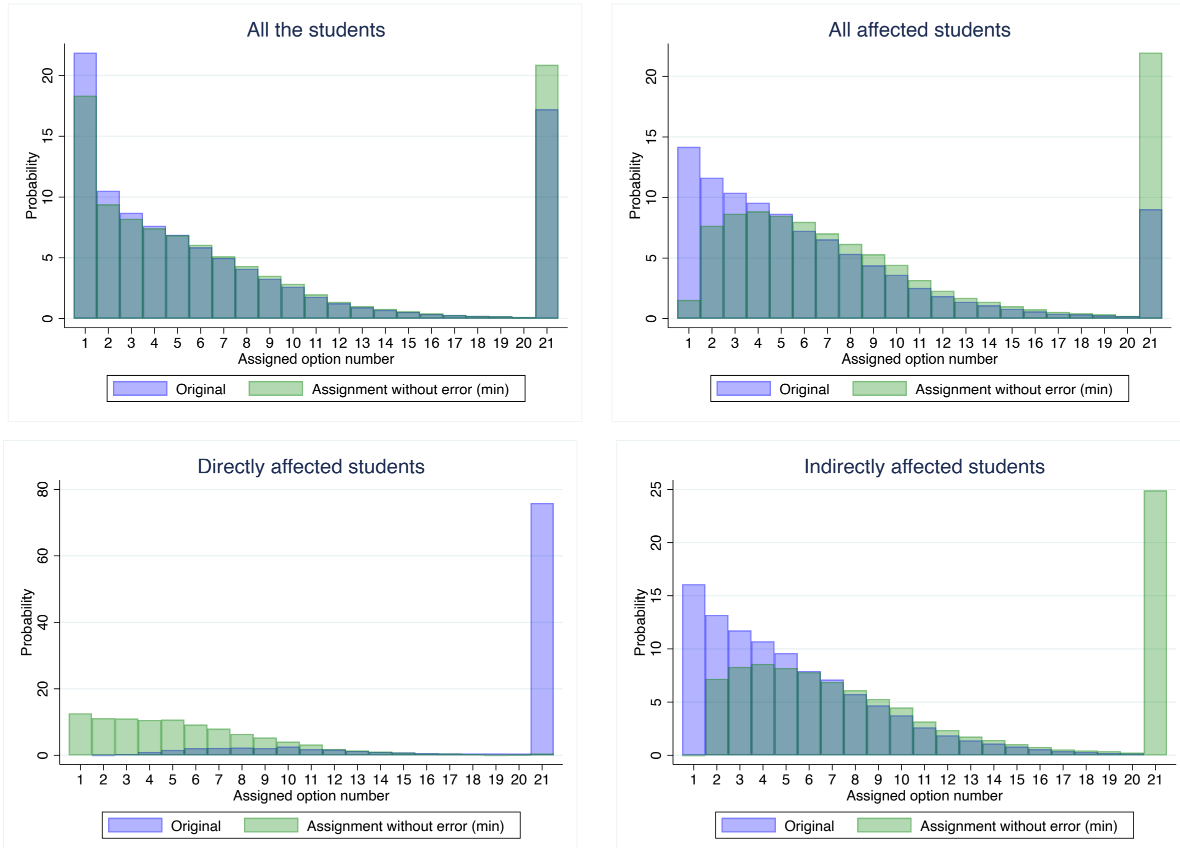


(b) SE

Notes: Panel 12a shows the distance's coefficients estimated of the 32 groups generated using the indicator variables man, moms and fathers' middle or less education, and no car. Panel 12b shows the standard errors of each of the 32 estimated distance's coefficients.

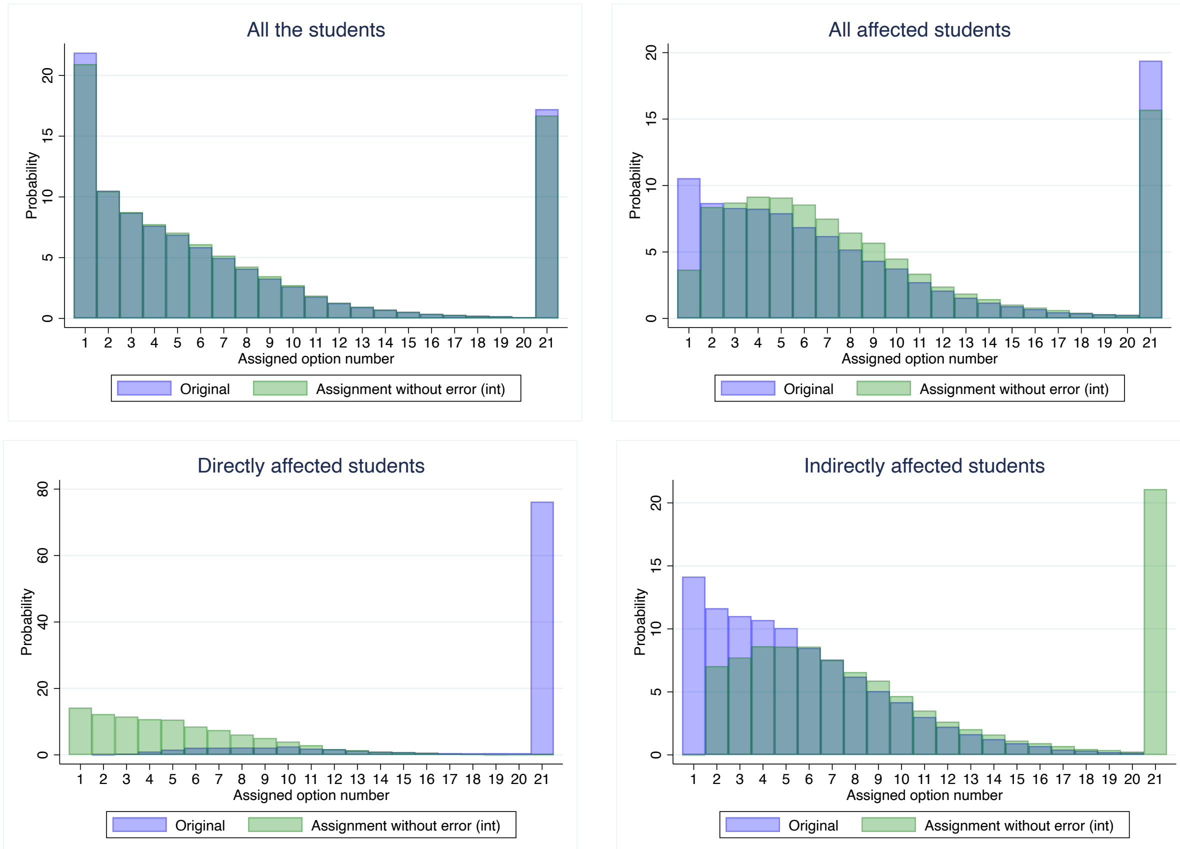
D Distribution of number of assigned school option

Figure 13: Original vs. Assignment without error (unfavorable)



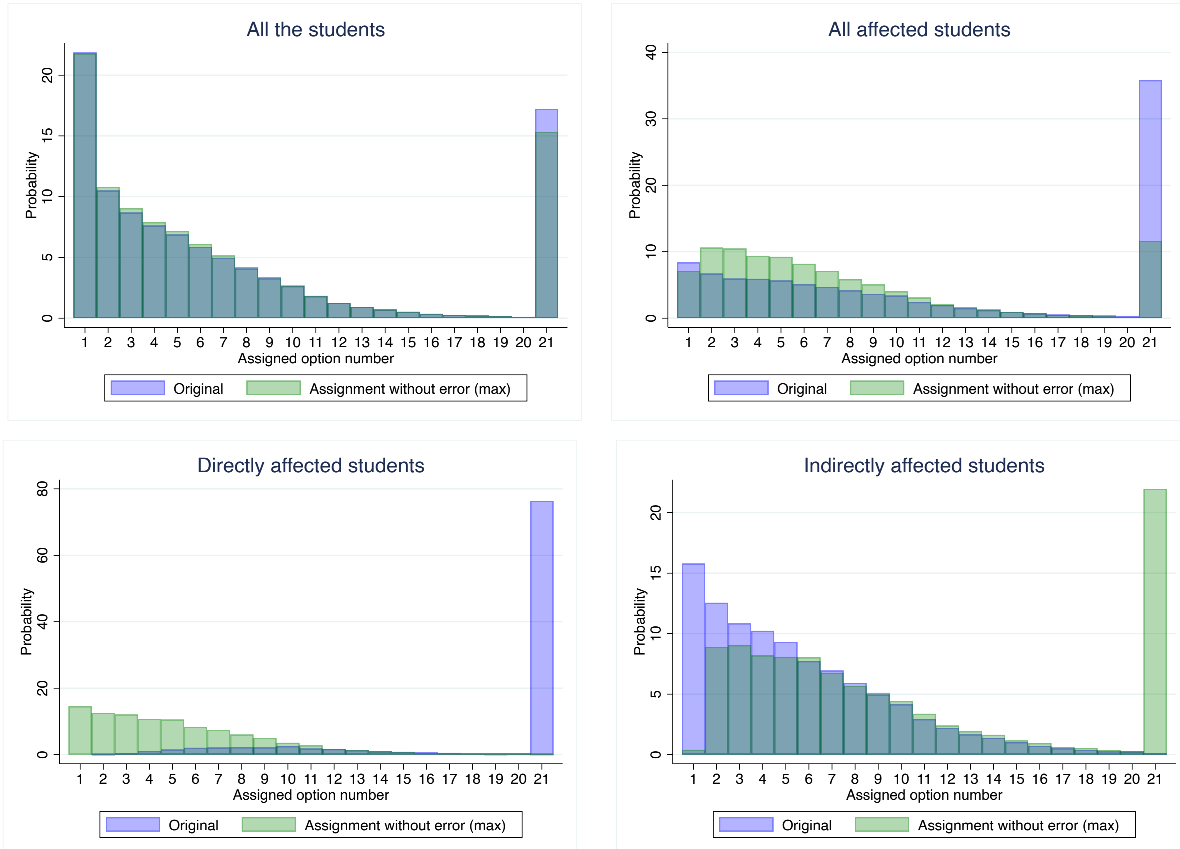
Notes: The figure shows how the number of assigned option's distribution changes between scenarios. The option number 21 indicates that the student was unmatched. Directly: students that had an incorrect grade and received a original assignment different to the assignment without error; indirectly: students that had a correct exam score and received an original assignment different to the assignment without error; affected: all the students affected directly or indirectly.

Figure 14: Original vs. Assignment without error (intermediate)



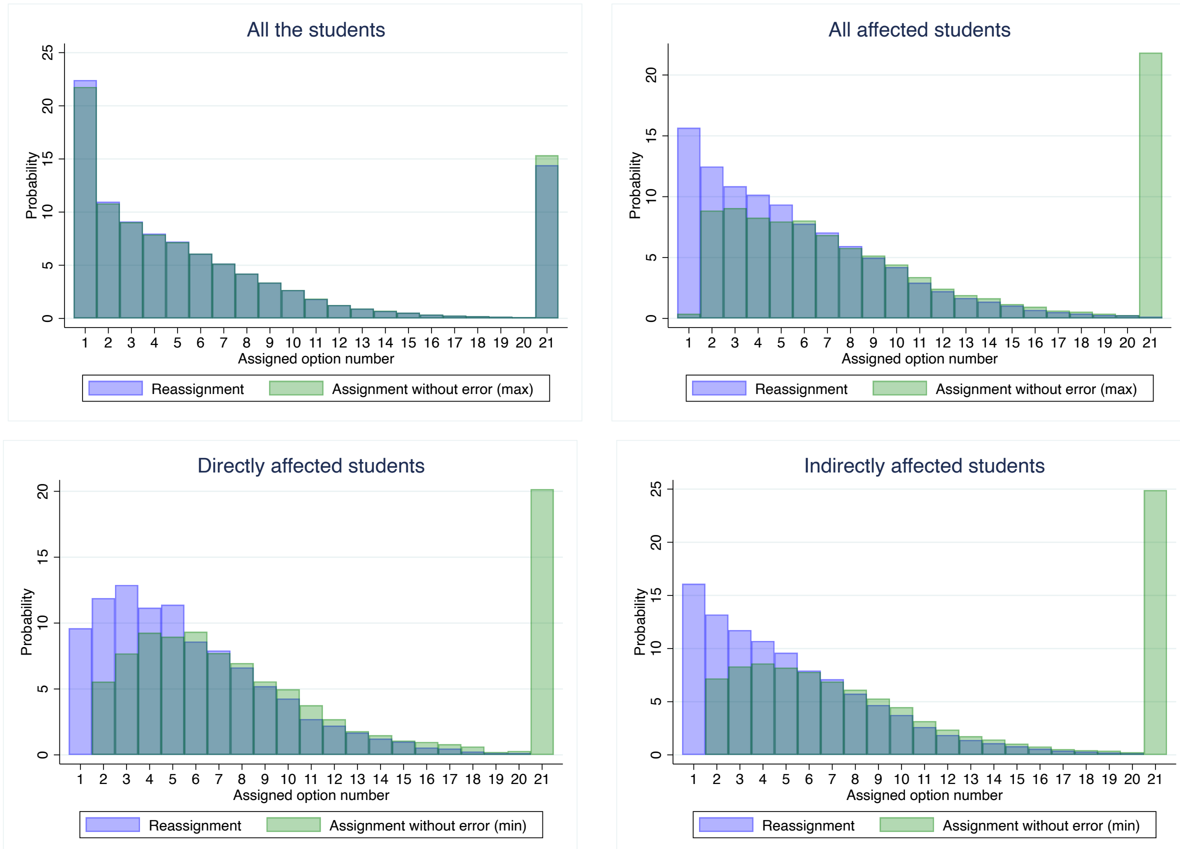
Notes: The figure shows how the number of assigned option's distribution changes between scenarios. The option number 21 indicates that the student was unmatched. Directly: students that had an incorrect grade and received a original assignment different to the assignment without error; indirectly: students that had a correct exam score and received an original assignment different to the assignment without error; affected: all the students affected directly or indirectly.

Figure 15: Original vs. Assignment without error (favorable)



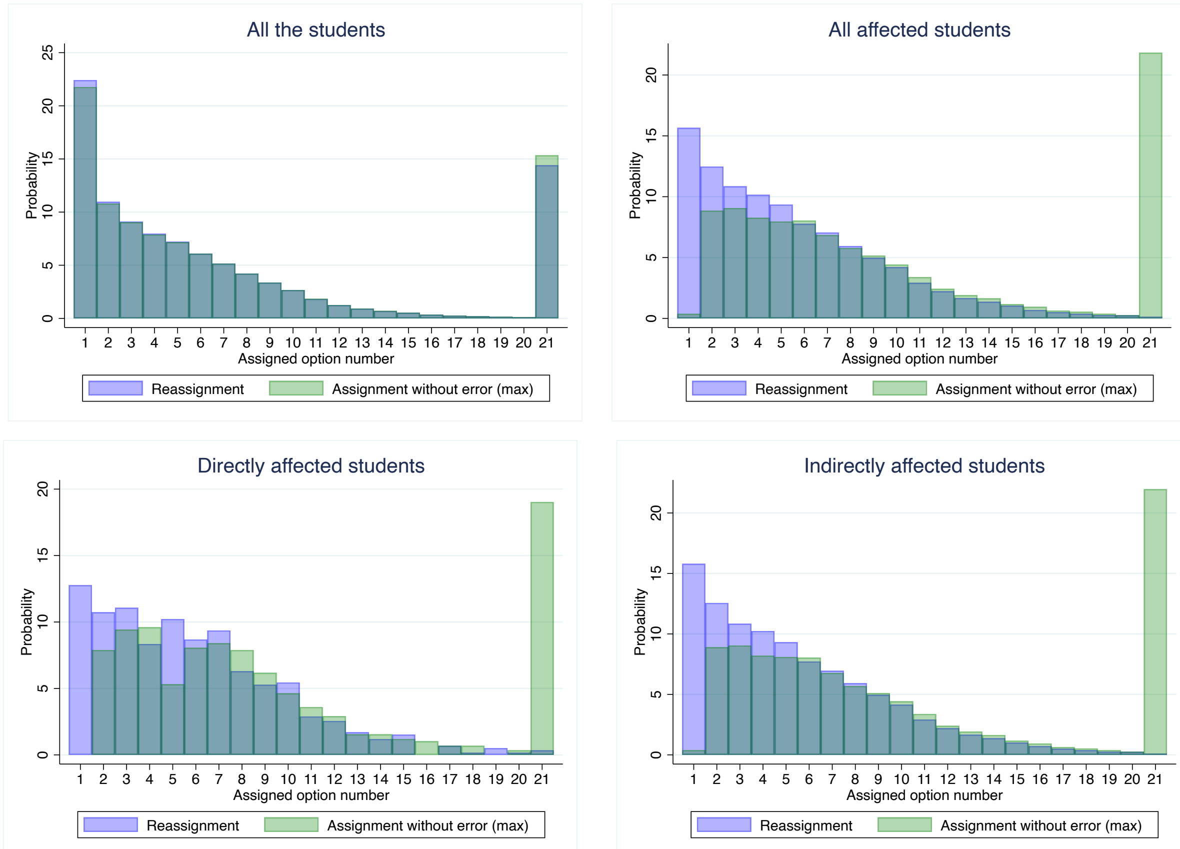
Notes: The figure shows how the number of assigned option's distribution changes between scenarios. The option number 21 indicates that the student was unmatched. Directly: students that had an incorrect grade and received a original assignment different to the assignment without error; indirectly: students that had a correct exam score and received an original assignment different to the assignment without error; affected: all the students affected directly or indirectly.

Figure 16: Reassignment vs. Assignment without error (unfavorable)



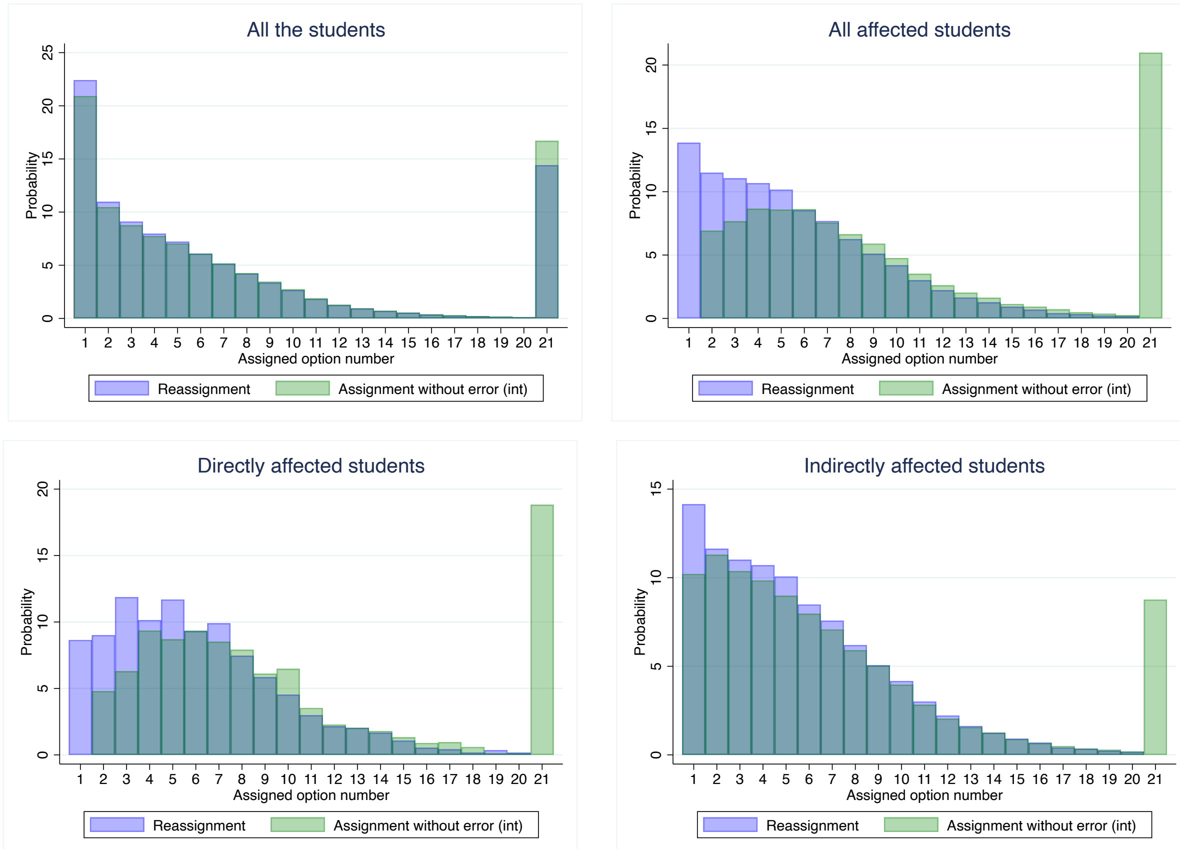
Notes: The figure shows how the number of assigned option's distribution changes between scenarios. The option number 21 indicates that the student was unmatched. Directly: students that had an incorrect grade and received a original assignment different to the assignment without error; indirectly: students that had a correct exam score and received an original assignment different to the assignment without error; affected: all the students affected directly or indirectly.

Figure 17: Reassignment vs. Assignment without error (favorable)



Notes: The figure shows how the number of assigned option's distribution changes between scenarios. The option number 21 indicates that the student was unmatched. Directly: students that had an incorrect grade and received a reassignment different to the assignment without error; indirectly: students that had a correct exam score and received an assignment different to the assignment without error; affected: all the students affected directly or indirectly.

Figure 18: Reassignment vs. Assignment without error (intermediate)

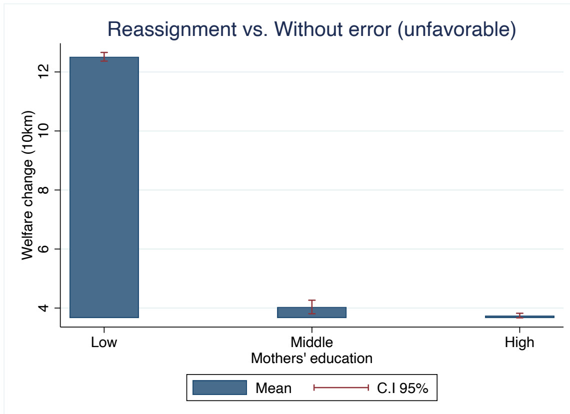


Notes: The figure shows how the number of assigned option's distribution changes between scenarios. The option number 21 indicates that the student was unmatched. Directly: students that had an incorrect grade and received a reassignment different to the assignment without error; indirectly: students that had a correct exam score and received an assignment different to the assignment without error; affected: all the students affected directly or indirectly.

E Welfare change graphs

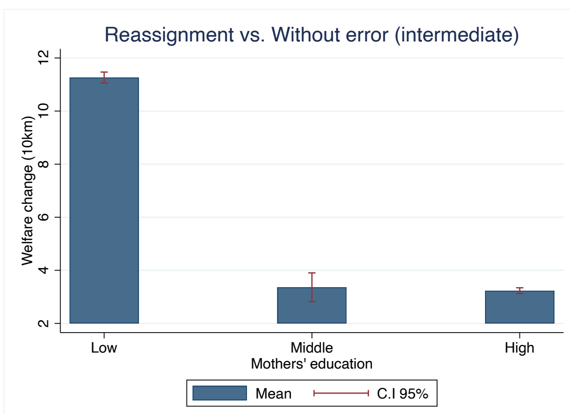
E.1 Welfare's change by socioeconomic group

Figure 19: Welfare's change distribution: Reassignment vs. Simulated (unfavorable)



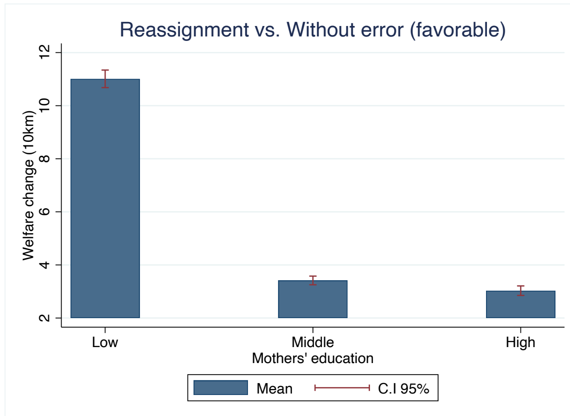
Notes: This figure shows the distribution of the welfare changes caused by the error. We show the results by level of the mother's education to point out a negative correlation between mean welfare change and the mother's education, which is a proxy for socioeconomic group.

Figure 20: Welfare's change distribution: Reassignment vs. Simulated (int)



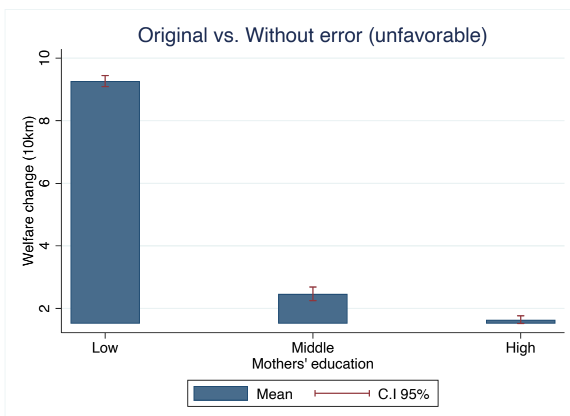
Notes: This figure shows the distribution of the welfare changes caused by the error. We show the results by level of the mother's education to point out a negative correlation between mean welfare change and the mother's education, which is a proxy for socioeconomic group.

Figure 21: Welfare's change distribution: Reassignment vs. Simulated (max)



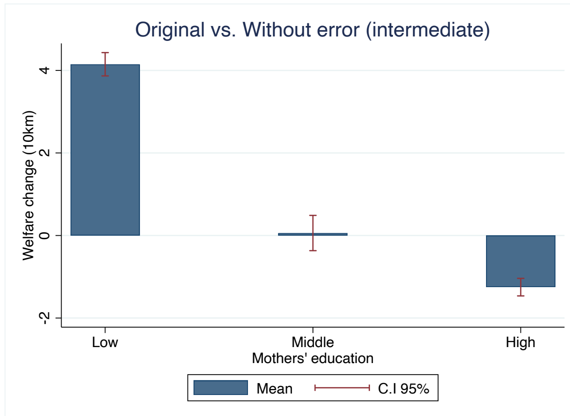
Notes: This figure shows the distribution of the welfare changes caused by the error. We show the results by level of the mother's education to point out a negative correlation between mean welfare change and the mother's education, which is a proxy for socioeconomic group.

Figure 22: Welfare's change distribution: Original vs. Without error (unfavorable)



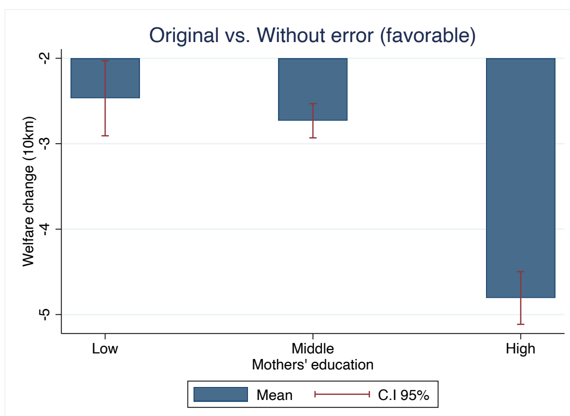
Notes: This figure shows the distribution of the welfare changes if the error had not been corrected. We show the results by level of the mother's education to point out a negative correlation between mean welfare change and the mother's education, which is a proxy for socioeconomic group.

Figure 23: Welfare's change distribution: Original vs. Without error (intermediate)



Notes: This figure shows the distribution of the welfare changes if the error had not been corrected. We show the results by level of the mother's education to point out a negative correlation between mean welfare change and the mother's education, which is a proxy for socioeconomic group.

Figure 24: Welfare's change distribution: Original vs. Without error (favorable)



Notes: This figure shows the distribution of the welfare changes if the error had not been corrected. We show the results by level of the mother's education to point out a negative correlation between mean welfare change and the mother's education, which is a proxy for socioeconomic group.

F Why low income benefits more?

Table 11: Percentage of indirectly affected applicants assigned to their first option

	Unfavorable		Intermediate		Favorable	
	Original	Without Error	Original	Without Error	Original	Without Error
Low	17.27%	3.89%	15.63%	3.96%	16.66%	4.33%
Mid	14.71%	2.6%	12.76%	2.78%	14.06%	3.19%
High	16%	1.75%	12.62%	2.2%	17.47%	2.06%

Notes: This table shows the percentage of indirectly affected applicants that were assigned to their first option by simulation. We have multiple columns for the original assignment because under each simulation the number of affected students changes, since it depends on the assignments without error.

Table 12: Percentage of directly affected applicants assigned to their first option

	Unfavorable		Intermediate		Favorable	
	Original	Without Error	Original	Without Error	Original	Without Error
Low	0%	8.46%	0%	9.49%	0%	10.04%
Mid	0%	10.33%	0%	11.82%	0%	12.54%
High	0%	19.82%	0%	22.71%	0%	23.3%

Notes: This table shows the percentage of directly affected applicants that were assigned to their first option by simulation. We have multiple columns for the original assignment because under each simulation the number of affected students changes, since it depends on the assignments without error.

Table 13: Percentage of directly affected applicants assigned

	Unfavorable			Intermediate			Favorable		
	N	Original (%)	Without Error (%)	N	Original (%)	Without Error (%)	N	Original (%)	Without Error (%)
Low	4,499	31.21	99.2	4,754	30.67	99.75	4,871	30.45	99.88
Mid	3,786	22.64	99.63	4,010	22.37	99.88	4,100	22.2	99.98
High	1,816	15.69	99.78	1,885	15.38	99.95	1,914	15.31	99.95
Total	10,101			10,649			10,885		

Notes: This table shows the percentage of directly affected applicants that were assigned to a educational option under each scenario by their mothers' education (low, mid, and high). The column N shows how many applicants belong to each of the categories under each simulation assumption.

Table 14: Percentage of indirectly affected applicants assigned

	Unfavorable			Intermediate			Favorable		
	N	Original (%)	Without Error (%)	N	Original (%)	Without Error (%)	N	Original (%)	Without Error (%)
Low	39,815	100	73.18	16,918	100	78.01	7,015	99.89	77.62
Mid	25,196	100	77.41	10,452	100	80.22	3,975	99.90	78.44
High	9,759	100	81.47	3,781	100	82.84	1,362	99.78	83.48
Total	74,770			31,151			12,352		

Notes: This table shows the percentage of indirectly affected applicants that were assigned to a educational option under each scenario by their mothers' education (low, mid, and high). The column N shows how many applicants belong to each of the categories under each simulation assumption.

Table 15: Mean number of option assigned of directly affected students assigned

	Unfavorable		Intermediate		Favorable	
	Original	Without Error	Original	Without Error	Original	Without Error
Low	9.51	5.72	9.58	5.52	9.59	5.44
Mid	10.32	5.75	10.28	5.49	10.25	5.41
High	11.21	4.67	11.23	4.35	11.22	4.25

Notes: This tables shows the mean of the number of option assigned for directly affected students conditional in been accepted.

Table 16: Mean number of option assigned of indirectly affected students assigned

	Unfavorable		Intermediate		Favorable	
	Original	Without Error	Original	Without Error	Original	Without Error
Low	5.05	6.75	5.32	6.87	5.31	6.74
Mid	5.39	7.04	5.75	7.26	5.75	7.04
High	4.93	6.64	5.42	6.98	4.95	6.23

Notes: This tables shows the mean of the number of option assigned for indirectly affected students conditional in been accepted.

Table 17: Proportion of Elite Schools in rank-ordered lists by socioeconomic level

Socioeconomic Level	Mean proportion
Low	35.49%
Mid	50.69%
High	65.62%

Notes: For Mexico City, the public high schools that are considered elite schools are those affiliated to UNAM or IPN, two of the most prestigious public universities of Mexico. Each row shows the proportion of the preferences submitted by the applicants that belong to these two elite subsystems.

G Simulated error

Table 18: Match outcomes of simulated scenarios with and without error.

	Unfavorable		Intermediate		Favorable	
	Error	Without Error	Error	Without error	Error	Without error
Num. of matched students	242,487	242,706	252,513	254,933	257,059	258,937
Num. of unmatched students	61,876	61,657	51,850	49,430	47,304	45,426
Mean num. of option assigned	8.11	4.84	7.44	4.61	7.11	4.52
% of students assigned first option	19.53%	18.79%	21.65%	21.39%	22.61%	22.25%

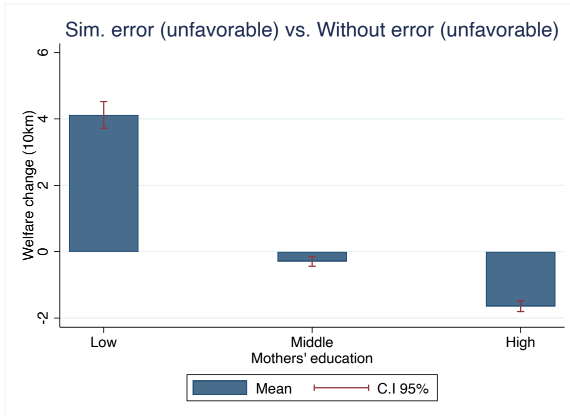
Notes: XXXX. The simulated match outcomes depend on the assumptions that we use about the schools' tie decisions. The percentage row is with respect the total number of eligible students (304,361).

Table 19: Affected students of simulated error

Simulation	Directly	Indirectly	Total affected	% of total
Unfavorable	11,141	29,634	40,775	13.39%
Intermediate	10,867	39,121	49,988	16.42%
Favorable	10,754	36,750	47504	15.60%

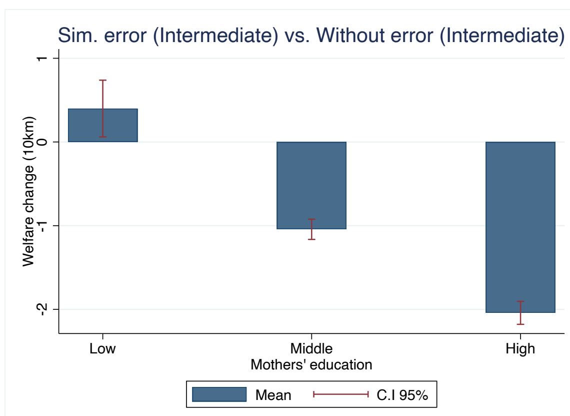
Notes: This table shows the number of affected students under the different simulations. An affected student is a participant that received a different assignment compared to the "universe" without error. Furthermore, a directly affected students are those that were affected because they received an incorrect score, while indirectly affected students are those that received a different match outcome due to spill-over effects. The percentage column is with respect of the total of eligible students (304,361).

Figure 25: Welfare's change distribution: Simulated error vs. Simulated without error (Unfavorable)



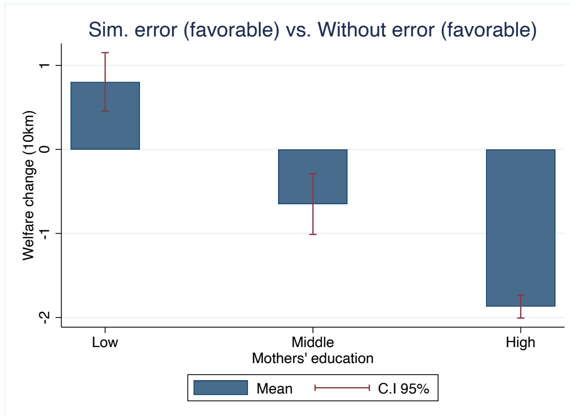
Notes: This figure shows the distribution of the welfare changes caused by the simulated error under the unfavorable assumption. We show the results by level of the mother's education to point out a negative correlation between mean welfare change and the mother's education, which is a proxy for socioeconomic group.

Figure 26: Welfare's change distribution: Simulated error vs. Simulated without error (Intermediate)



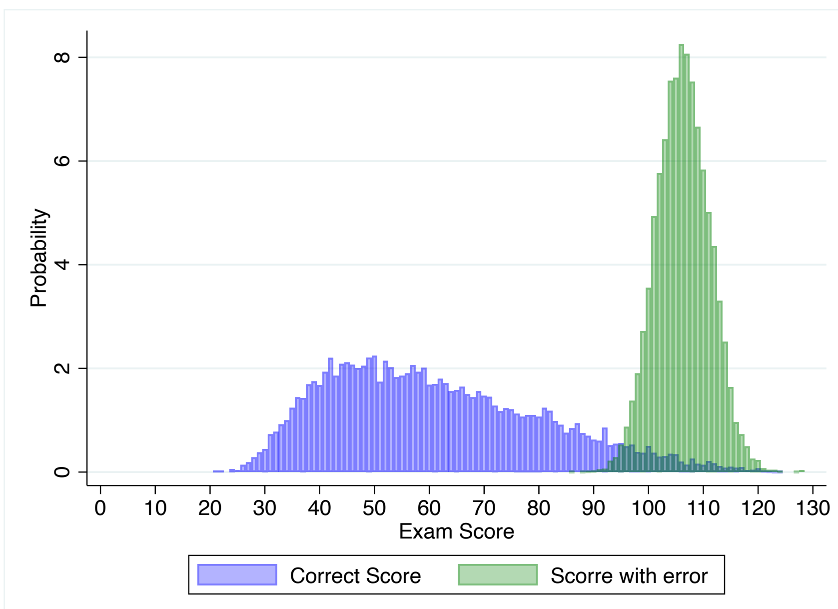
Notes: This figure shows the distribution of the welfare changes caused by the simulated error under the intermediate assumption. We show the results by level of the mother's education to point out a negative correlation between mean welfare change and the mother's education, which is a proxy for socioeconomic group.

Figure 27: Welfare's change distribution: Simulated error vs. Simulated without error (favorable)



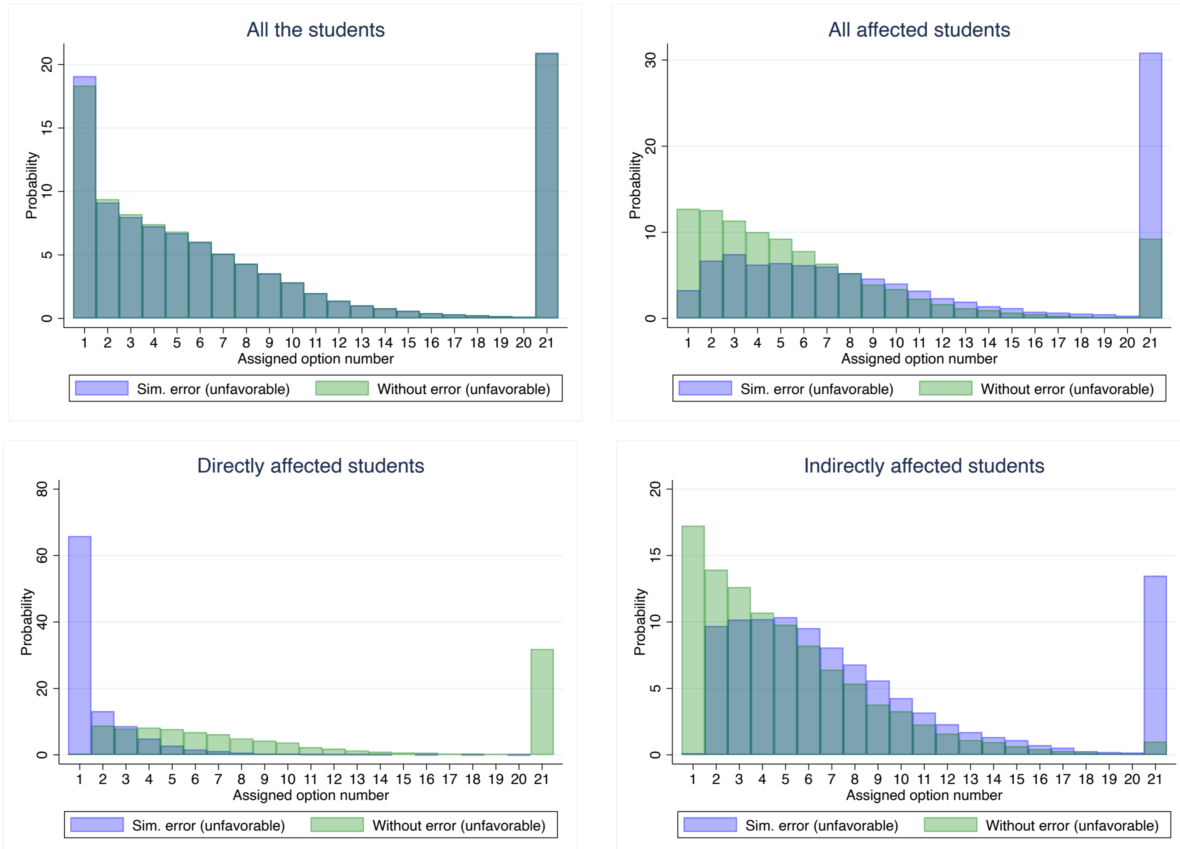
Notes: This figure shows the distribution of the welfare changes caused by the simulated error under the favorable assumption. We show the results by level of the mother's education to point out a negative correlation between mean welfare change and the mother's education, which is a proxy for socioeconomic group.

Figure 28: Exam scores' distribution for the students affected by the simulated error



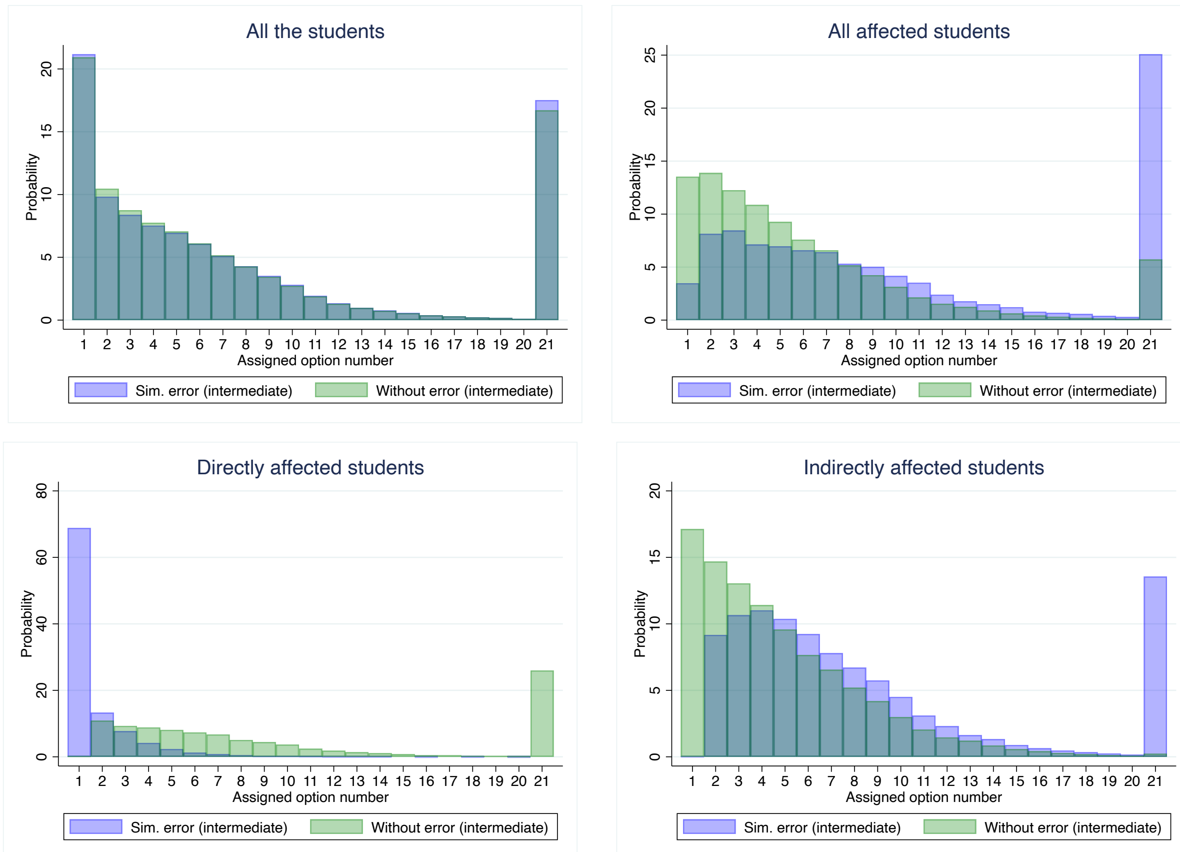
Note: This graph shows how the exam scores' distribution change when we simulated an error, compare to the empirical error, on the other side of the exam score's distribution. The total number of affected applicants by the simulated error is 13,712. The first score contains the error, while the second exam score is without error.

Figure 29: Sim. Error (unfavorable) vs. Assignment without error (unfavorable)



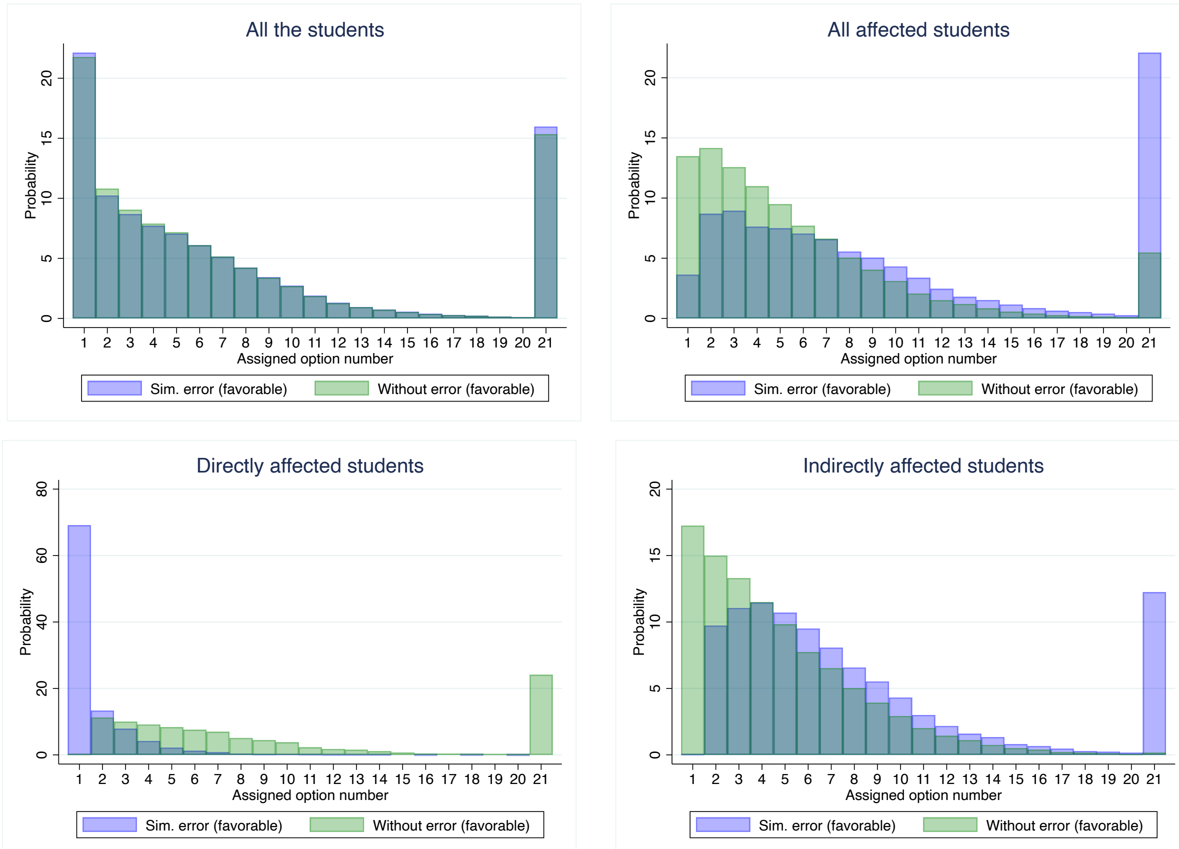
Notes: The figure shows how the number of assigned option's distribution changes between scenarios. The option number 21 indicates that the student was unmatched. Directly: students that had an incorrect grade and received an assignment different to the assignment without error; indirectly: students that had a correct exam score and received an assignment different to the assignment without error; affected: all the students affected directly or indirectly.

Figure 30: Sim. Error (intermediate) vs. Assignment without error (intermediate)



Notes: The figure shows how the number of assigned option's distribution changes between scenarios. The option number 21 indicates that the student was unmatched. Directly: students that had an incorrect grade and received an assignment different to the assignment without error; indirectly: students that had a correct exam score and received an assignment different to the assignment without error; affected: all the students affected directly or indirectly.

Figure 31: Sim. Error (favorable) vs. Assignment without error (favorable)



Notes: The figure shows how the number of assigned option's distribution changes between scenarios. The option number 21 indicates that the student was unmatched. Directly: students that had an incorrect grade and received an assignment different to the assignment without error; indirectly: students that had a correct exam score and received an assignment different to the assignment without error; affected: all the students affected directly or indirectly.

Table 20: Percentage of directly affected applicants assigned to their first option

	Unfavorable		Intermediate		Favorable	
	Sim. Error	Without Error	Sim. Error	Without Error	Sim. Error	Without Error
Low	5.03%	70.99%	2.21%	73.51%	0.91%	73.82%
Mid	3.74%	60.39%	1.72%	63.97%	0.87%	64.17%
High	5.47%	54.64%	3.11%	58.12%	1.93%	58.43%

Notes: This table shows the percentage of directly affected applicants that were assigned to their first option by simulation. We have multiple columns for the original assignment because under each simulation the number of affected students changes, since it depends on the assignments without error.

Table 21: Percentage of indirectly affected applicants assigned to their first option

	Unfavorable		Intermediate		Favorable	
	Sim. Error	Without Error	Sim. Error	Without Error	Sim. Error	Without Error
Low	23.83%	.32%	19.47%	0.11%	17.61%	0.09%
Mid	21.18%	0.05%	17.14%	0.01%	16.06%	0.02%
High	22.33%	0%	17.52%	0%	16.2%	0%

Notes: This table shows the percentage of affected applicants that were assigned to their first option by simulation. We have multiple columns for the original assignment because under each simulation the number of affected students changes, since it depends on the assignments without error.

Table 22: Percentage of directly affected applicants assigned

	Unfavorable			Intermediate			Favorable		
	N	Sim. Error (%)	Without Error (%)	N	Sim. Error (%)	Without Error (%)	N	Sim. Error (%)	Without Error (%)
Low	6,563	78.84	69.71	6,383	78.23	75.86	6,309	77.98	77.95
Mid	3,555	75.19	68.27	3,486	74.56	73.61	3,461	74.34	75.15
High	1,023	72.92	68.52	998	72.24	72.24	984	71.65	73.17
Total	11,141			10,867			10,754		

Notes: This table shows the percentage of directly affected applicants that were assigned to a educational option under each scenario by their mothers' education (low, mid, and high). The column N shows how many applicants belong to each of the categories under each simulation assumption.

Table 23: Percentage of indirectly affected applicants assigned

	Unfavorable			Intermediate			Favorable		
	N	Sim. Error (%)	Without Error (%)	N	Sim. Error (%)	Without Error (%)	N	Sim. Error (%)	Without Error (%)
Low	12,398	95.56	98.31	16,906	95.78	99.59	15,813	95.45	99.71
Mid	11,383	93.98	99.31	14,762	94.30	99.88	13,880	94.5	99.89
High	5,853	93.64	99.90	7,453	93.75	99.95	7,057	93.69	99.98
Total	29,634			39,121			36,750		

Notes: This table shows the percentage of indirectly affected applicants that were assigned to a educational option under each scenario by their mothers' education (low, mid, and high). The column N shows how many applicants belong to each of the categories under each simulation assumption.

Table 24: Mean number of option assigned of directly affected students assigned

	Unfavorable		Intermediate		Favorable	
	Sim. Error	Without Error	Sim. Error	Without Error	Sim. Error	Without Error
Low	1.73	6.07	1.62	5.89	1.61	5.83
Mid	2.04	6.71	1.9	6.55	1.86	6.45
High	2.21	6.63	2.01	6.38	1.99	6.29

Notes: This tables shows the mean of the number of option assigned for directly affected students conditional in been accepted.

Table 25: Mean number of option assigned of indirectly affected students assigned

	Unfavorable		Intermediate		Favorable	
	Sim. Error	Without Error	Sim. Error	Without Error	Sim. Error	Without Error
Low	6.33	4.79	6.36	4.75	6.3	4.71
Mid	6.62	4.99	6.53	4.91	6.38	4.78
High	6.32	4.64	6.28	4.56	6.14	4.44

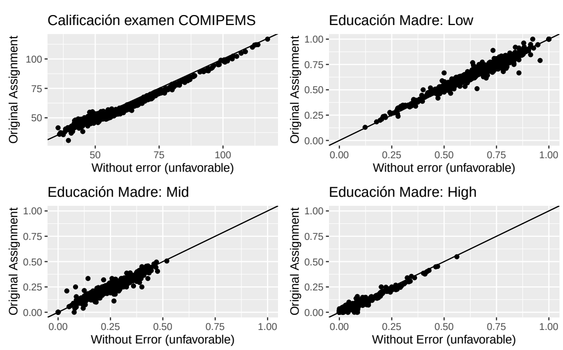
Notes: This tables shows the mean of the number of option assigned for indirectly affected students conditional in been accepted.

H Outcomes

	(1) FE Promedio	(2) FE Prom. Mat.	(3) FE Desercion	(4) FE indisciplina	(5) FE Inasistencia Ago	(6) FE Inasistencia Ene
Examen Comipems	0.137*	0.094	-0.025	0.003	-6.542***	-6.589***
	(0.076)	(0.095)	(0.020)	(0.026)	(1.258)	(1.204)
Constant	8.260***	7.809***	0.076***	0.099***	8.835***	10.190***
	(0.055)	(0.069)	(0.014)	(0.019)	(0.903)	(0.876)
N	213.0	221.0	232.0	223.0	241.0	244.0
R^2	0.02	0.00	0.01	0.00	0.10	0.11

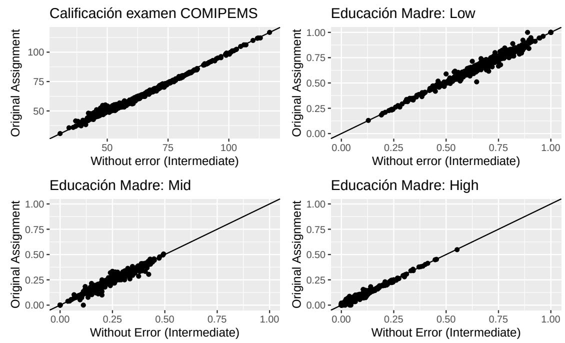
I Peers effects

Figure 32: Comparison of peer effects under different scenarios



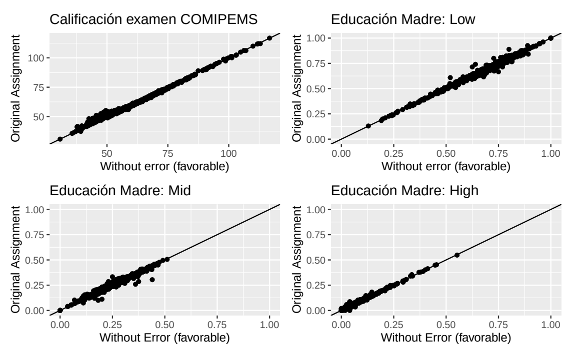
Notes: Each dot represents the 700's schools average sociodemographic composition. A 45 degrees line is represented in the graph for comparisons. A dot above or under the 45 degrees line indicates that the school has a different sociodemographic composition under the two scenarios compared.

Figure 33: Comparison of peer effects under different scenarios



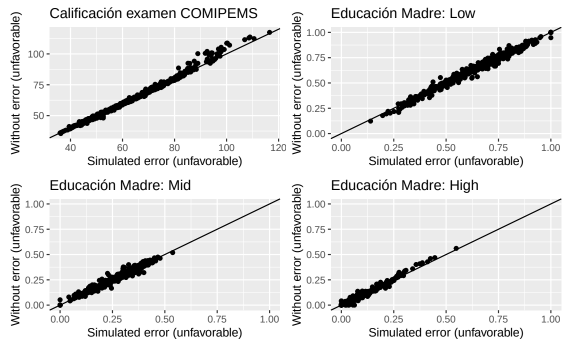
Notes: Each dot represents the 700's schools average sociodemographic composition. A 45 degrees line is represented in the graph for comparisons. A dot above or under the 45 degrees line indicates that the school has a different sociodemographic composition under the two scenarios compared.

Figure 34: Comparison of peer effects under different scenarios



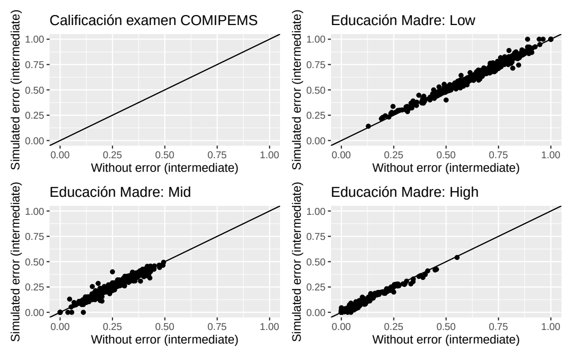
Notes: Each dot represents the 700's schools average sociodemographic composition. A 45 degrees line is represented in the graph for comparisons. A dot above or under the 45 degrees line indicates that the school has a different sociodemographic composition under the two scenarios compared.

Figure 35: Comparison of peer effects under different scenarios



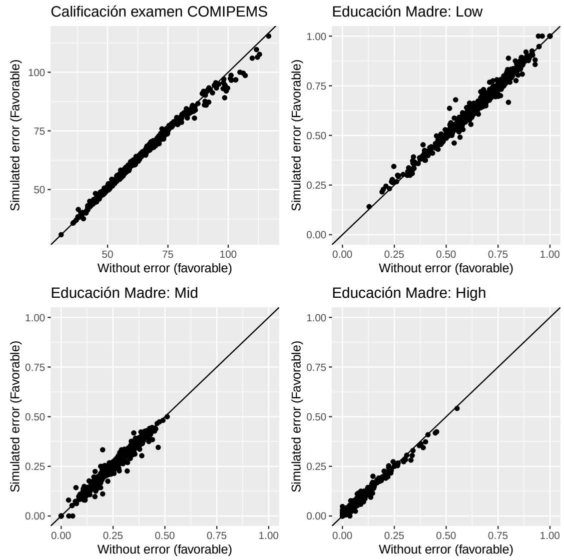
Notes: Each dot represents the 700's schools average sociodemographic composition. A 45 degrees line is represented in the graph for comparisons. A dot above or under the 45 degrees line indicates that the school has a different sociodemographic composition under the two scenarios compared.

Figure 36: Comparison of peer effects under different scenarios



Notes: Each dot represents the 700's schools average sociodemographic composition. A 45 degrees line is represented in the graph for comparisons. A dot above or under the 45 degrees line indicates that the school has a different sociodemographic composition under the two scenarios compared.

Figure 37: Comparison of peer effects under different scenarios



Notes: Each dot represents the 700's schools average sociodemographic composition. A 45 degrees line is represented in the graph for comparisons. A dot above or under the 45 degrees line indicates that the school has a different sociodemographic composition under the two scenarios compared.

J Outcomes Econometrics

Table 26: FE regression using dummys

	(1) FE Promedio	(2) FE Prom. Mat.	(3) FE Desercion	(4) FE indisciplina	(5) FE Inasistencia Ago	(6) FE Inasistencia Ene
Examen Comipems	0.229*	0.089	-0.050	0.064*	-7.681***	-9.691
	(0.132)	(0.167)	(0.086)	(0.034)	(2.412)	(6.652)
Constant	11668.405***	16297.597***	-109.779***	2340.351***	-4190.704***	-1.13e+04***
	(0.095)	(0.121)	(0.062)	(0.025)	(1.740)	(4.832)
N	214	211	235	210	244	243
R^2	0.01	0.00	0.00	0.02	0.04	0.01

Table 27: FE regression no dummys

	(1) FE Promedio	(2) FE Prom. Mat.	(3) FE Desercion	(4) FE indisciplina	(5) FE Inasistencia Ago	(6) FE Inasistencia Ene
Examen Comipems	0.161*	0.034	-0.041**	0.015	-6.277***	-5.998***
	(0.082)	(0.106)	(0.017)	(0.026)	(1.208)	(1.218)
Constant	8.293***	7.848***	0.070***	0.091***	8.751***	9.755***
	(0.060)	(0.077)	(0.013)	(0.019)	(0.869)	(0.886)
N	216.0	215.0	235.0	207.0	239.0	234.0
R^2	0.02	0.00	0.02	0.00	0.10	0.09

Table 28: FE regression with IVs

	(1) FE Promedio	(2) FE Prom. Mat.	(3) FE Desercion	(4) FE indisciplina	(5) FE Inasistencia Ago	(6) FE Inasistencia Ene
Examen Comipems	0.259	0.182	0.017	0.082	-12.107***	-8.685**
	(0.256)	(0.325)	(0.053)	(0.078)	(3.772)	(3.624)
Class size	-0.000	-0.000	-0.000	-0.000	0.002	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.003)
Constant	8.437***	8.023***	0.109***	0.129**	5.629**	8.807***
	(0.183)	(0.229)	(0.038)	(0.057)	(2.685)	(2.595)
N	216	215	235	207	239	234
R^2	0.04	0.02	.	.	0.02	0.07