Peer Effects in Matching Systems and the Effects of Immigrant Classmates on Native Students

Mohit Karnani*
MIT, Harvard

November 6, 2026

Abstract

This paper studies the estimation of peer effects in centralized school choice systems. These centralized assignment mechanisms often ration seats by lottery, and thus classroom composition and the presence of certain types of peers in a cohort can be the result of a random assignment process. We show how this random variation can be exploited to identify peer effects. The estimator relies on two different propensity scores that can be extracted from a general class of centralized matching mechanisms, and it emulates the setting of a stratified cluster-randomized experiment. We use this approach to study the impact of immigrant classmates on native first graders in Chile's public school system. Our findings suggest that the presence of immigrant classmates has no overall significant impact on native children's academic achievement, but it does reduce reported victimization by classmates and boosts reports of other positive attitudes. Academic and behavioral gains are concentrated among cohorts that are formed from scratch, i.e. where the entire cohort is new to the school. Additionally, we find that incumbent students already enrolled in a school whose grade receives a migrant entrant are more likely to remain enrolled in that school. These results provide insights into the dynamics of peer effects in educational settings and have implications for policy design in school assignment systems.

*Email: karnani@mit.edu. I thank the Chilean Ministry of Education and the Agencia de Calidad de la Educación for access to the data used in this paper.

1 Introduction

School districts around the world assign seats through centralized mechanisms that ration oversubscribed schools by lottery. Deferred acceptance and related mechanisms, introduced by Gale and Shapley (1962) and adapted to school choice by Abdulkadiroğlu and Sönmez (2003), now allocate seats in Boston, New York, Denver, Amsterdam and many other cities, and since the late 2010s they have assigned every student who applies to a Chilean school that receives public funds.

The random tiebreakers of these assignment mechanisms have been used to estimate causal effects within centralized school choice systems. Abdulkadiroğlu et al. (2017) show that the lotteries embedded in a wide class of these mechanisms generate a stratified randomized trial of school assignment, whose strata are summarized by a propensity score that the mechanism itself determines. Research designs built on this insight have been used to estimate causal treatment effects of schools on students, such as value added effects of charter schools or schools rated highly by district report cards (Abdulkadiroğlu et al., 2017, 2022).

A school, however, is also a set of classmates, and questions about who sits beside whom are at the center of debates over tracking, desegregation, school choice and immigration.[1] Families choose schools, schools select students, and students influence one another, so the association between a student's outcomes and the characteristics of her peers mixes the effect of those peers with the reasons they came to be together (Manski, 1993; Angrist, 2014). Most credible estimates therefore rest on settings in which peers are assigned at random, such as college roommates (Sacerdote, 2001) and squadrons at a military academy (Carrell et al., 2013), on the idiosyncratic variation in cohort composition that arises within schools over time (Hoxby, 2000), or on the admission cutoffs of selective schools, across which the peers a student meets change discontinuously (Abdulkadiroğlu et al., 2014).

This paper shows how the lotteries embedded in centralized school assignment can be used to estimate peer effects. When a school-grade cohort is oversubscribed, the draw that seats one applicant rather than another determines the composition of the cohort, so the lottery that generates the experiment in school assignment generates a second experiment in peer composition. We study this second experiment and use it to estimate the effects of immigrant classmates on native children in Chile.

The treatment in our peer-effects design is an attribute of the assignment unit (school-grade) rather than of the student. A native applicant's exposure to an immigrant classmate depends both on her own assignment and on the assignments of the immigrant applicants. We show that conditioning on natives' probability of exposure to immigrant classmates, the analog of the assignment propensity score (Abdulkadiroğlu et al., 2017), balances natives' attributes but not the attributes of the cohorts natives are assigned to. Exposed applicants are disproportionately those who landed in the cohorts immigrant families rank, and such cohorts may differ from others in size, location and composition. A comparison that controls for the exposure score alone, or recenters exposure at it (Borusyak and Hull, 2023), therefore estimates a peer effect plus a school effect.

Our solution is to recenter each applicant's realized exposure at its cohort's expectation given her own assignment. This modified recentered exposure is orthogonal to

[1] See Epple and Romano, 2011 for a review of the peer effects literature.

everything fixed before the lottery and to the identity and quality of the school the applicant is assigned to. It is a recentered instrument in the sense of Borusyak and Hull (2023), with the recentering conditional on own assignment, and it identifies a weighted average of peer effects.

The same lotteries also run another simpler experiment on incumbent students already enrolled in a school, whose cohort the lottery does not choose. For these incumbents the analysis is a stratified cluster-randomized trial with assignment probabilities the mechanism delivers. To the best of our knowledge, ours is the first analysis to derive the propensity scores that a centralized assignment mechanism implies for peer exposure and to use them to estimate peer effects.

Our test bed is Chile's *Sistema de Admisión Escolar* (SAE), a national deferred-acceptance match with lottery tie-breaking. Foreign-born enrollment in Chilean schools grew from 65,764 students in 2016 to 242,933 in 2022, with students recorded as Venezuelan accounting for 48 percent of the increase. We study the native children who applied for a first-grade seat through the 2018 and 2019 admission processes, together with the native children already enrolled in schools that could receive foreign-born students.

First grade is the grade in which primary-school classes are formed and tend to persist.[2] The children of both processes took the national SIMCE assessment three years after entry, which measures math and reading proficiency, and the assessment also fields questionnaires to the students, which we use to measure non-academic impacts. In addition, first grade lets us compare two kinds of cohort. In a school without a kindergarten, every first grader is new and the match forms the cohort from scratch. In a school with kindergarten, the children of its kindergarten can continue into first grade without applying. Because whether a school has a kindergarten is fixed before the lottery, the design runs a separate experiment in each kind of school.

The empirical analysis begins with the design itself. Comparisons of exposed and unexposed natives that control for nothing, or for the exposure score alone, are unbalanced on the attributes of the cohorts natives are assigned to, while recentered comparisons are balanced on every attribute we measure, in expectation and in joint tests at the realized lottery. The simulated scores also predict the lottery that took place, i.e. grouped by either score, the share of native applicants whose realized cohort received a migrant applicant lies on the 45-degree line.

Our lottery-based estimates show no cost of immigrant classmates for native achievement. A native child whose first-grade cohort receives an immigrant applicant scores no differently in reading and mathematics three years later. Native children instead report being victimized by their classmates less, by 0.070 standard deviations from assignment and 0.191 from having an enrolled immigrant grade-mate. The decline appears whether or not the school has a kindergarten, but where the match forms the first-grade cohort from scratch the arrival of an immigrant applicant also boosts academic achievement. In newly formed cohorts, native children whose cohort receives a foreign-born classmate score 0.143 standard deviations higher in reading and mathematics and report feeling better at school, by 0.153 standard deviations, while in schools whose kindergarten children continue into first grade neither moves.

[2] In contrast to other grades, such as kindergarten or 8th grade, where many students must switch to a different school to attend 1st grade or 9th grade (1st year of high school).

Schools that teach first grade in a single class, where the class is the grade, suggest that the decline in victimization runs through classmates (as opposed to grademates). There, the grade-level and class-level first stages nearly coincide, and the estimated effect of an immigrant classmate is almost double than in schools with several classes.[3] The children already enrolled in the schools immigrants enter do not leave, i.e. we find no evidence of native flight. If anything, they are more likely to stay, by 2.2 percentage points a year after entry. Without the lottery, a naive comparison using the same data would suggest that immigrant classmates lower native attendance, by 1.2 percentage points of school days, a conclusion the design does not support. This suggests that selection bias in our setting is strong enough to reverse the sign of the peer effects of immigrant classmates on native outcomes under simple mean comparisons.

Our analysis connects several lines of work. The first is the design-based analysis of centralized assignment to estimate causal effects pioneered by Abdulkadiroğlu et al. (2017), which we extend from treatments that are attributes of a student's own assignment (e.g. assigned to a charter school or not) to treatments that are attributes of the unit she is assigned to (e.g. does the cohort receive a migrant student or not). The second is the use of recentered instruments when an exogenous shock reaches units through nonrandom exposure (Borusyak et al., 2022; Borusyak and Hull, 2023). In our setting the exposure is itself generated by the mechanism, and the recentering must condition on the student's own assignment, since recentering at her unconditional probability of exposure leaves the school effect in place. Third, we provide lottery-based estimates of the effects of immigrants in a system absorbing a large migration wave, mostly due to the Venezuelan exodus in the 2010s ().

The remainder of the paper is structured as follows. Section 2 describes the class of mechanisms of interest to us and the Chilean data. Section 3 develops the theory of the exposure scores and reports balancing tests. Section 4 estimates the effects of immigrant classmates on the children the lottery places and on those already enrolled, and Section 5 concludes.

2 Centralized Assignment and Peer Exposure

2.1 Conditional Independence and ETE

We work in the standard centralized school choice environment (Abdulkadiroğlu et al., 2017), with one change of vocabulary: because the treatment of interest is an attribute of the unit to which a student is assigned, we call the units *cohorts* rather than schools. A cohort is a school-grade, which is what the Chilean mechanism assigns.[4] Let I denote a set of applicants, indexed by i, and let $g = 0, 1, \ldots, G$ index cohorts, where $g = 0$ represents an outside option, and each cohort $g \geq 1$ belongs to a school $s(g)$. Applicant i has strict preferences $\succ_i$ over cohorts and a priority ρ_{ig} at each, and her type is $\theta_i = (\succ_i, \rho_i)$. A mechanism is a set of rules that determines an assignment μ, a cohort μ_i

[3] In addition, in a placebo exercise using the same lotteries but analyzing the exposure of the kindergarten of the school a native is assigned to, we find no effects.

[4] This is to avoid ambiguity with migrants that can be assigned to the same school as a native, but in a different grade, in which case we do not consider them to be peers. This distinction is also useful because we later study the effects of grademates vs classmates.

for every applicant, as a function of preferences, priorities, capacities and a tie-breaking variable that is randomly drawn. We write $D_i(g) = \mathbf{1}\{\mu_i = g\}$ and

$$p_{ig} = \mathbb{P}[D_i(g) = 1] \tag{1}$$

for the probability that the mechanism assigns applicant i to cohort g. Throughout, probabilities and expectations are taken over the lottery, holding the population of applicants, their types and their potential outcomes fixed.

A mechanism satisfies equal treatment of equals (ETE) when applicants with the same type have the same assignment probability at every cohort, so that $p_{ig} = p_g(\theta_i)$. Say that an attribute W_i is *fixed under re-randomization* if its value is the same in every assignment the mechanism can produce; covariates are fixed in this sense, and so are potential outcomes. Abdulkadiroğlu et al. (2017) show that, for a mechanism satisfying ETE and any such W_i,

$$\mathbb{P}[D_i(g) = 1 \mid W_i = w, \theta_i = \theta] = \mathbb{P}[D_i(g) = 1 \mid \theta_i = \theta] \equiv p_g(\theta) \tag{2}$$

for every w taken by an applicant of type θ. The function $p_g(\theta)$ is the propensity score of assignment to cohort g for an applicant of type θ, and (2) implies that the mechanism runs a stratified randomized trial whose strata are types. Deferred acceptance with lottery tie-breaking, whether the lottery draws one number per applicant or one per applicant and school (as in Chile), is one of the most popular mechanisms in the ETE class.

2.2 Peer Exposure as a Treatment

Let $\mathcal{A} \subset I$ be a set of applicants whose presence in a cohort is the treatment of interest. In our application, $\mathcal{A}$ contains the applicants the Ministry's enrollment records identify as foreign-born, whom we call migrant applicants. Throughout, an immigrant grademate shares a native's school-grade cohort and a classmate her class. We speak of immigrant classmates where the distinction does not matter, but the main estimates concern grademates, as that is the unit of assignment determined by the SAE mechanism.[5] Membership in $\mathcal{A}$ does not respond to the lottery and the applicants in $\mathcal{N} \subseteq I \setminus \mathcal{A}$ are the population on which peer effects are estimated, who we call native applicants. For every cohort $g \geq 1$ define

$$N_g = \sum_{j \in \mathcal{A}} D_j(g), \qquad A_g = \mathbf{1}\{N_g \geq 1\}, \tag{3}$$

the number of migrant applicants the mechanism sends to cohort g and the indicator that it sends at least one, with $N_0 = A_0 = 0$. The *assigned exposure* of a native applicant is the exposure of the cohort she is assigned to, $Z_i = A_{\mu_i} = \sum_g D_i(g) A_g$.

A single lottery draw now runs two experiments at the same time. The first places each native applicant in a cohort, and the second places each migrant applicant and thereby determines A_g for every cohort. Exposure of this kind violates the no-interference part of the stable unit treatment value assumption (SUTVA), since a native's outcome depends on the assignments of other applicants. We restrict this dependence with an exposure mapping:

[5] The Chilean system assigns applicants to school-grade cohorts, and then schools determine how students are placed in classrooms within their grades. Section 4.2 separates the two.

Assumption 1 (Exposure Mapping). *For every $i \in \mathcal{N}$ there are numbers $\{Y_i(g,n) : g = 0,\dots,G;\ n = 0,1,\dots\}$, fixed under re-randomization, such that $Y_i = Y_i(\mu_i, N_{\mu_i})$ for every assignment the mechanism can produce, with $Y_i(0,n) = Y_i(0,0)$ for all n.*

This says that a native's outcome depends on the cohort she is assigned to and on how many migrant applicants it receives, and on nothing else about the assignment. Most of our results use the *binary mapping*, $Y_i(g,n) = Y_i(g,1)$ for all $n \geq 1$, under which the effect on i of a migrant peer at cohort g is $\Delta_i(g) = Y_i(g,1) - Y_i(g,0)$. The analog of the constant-effects model in Abdulkadiroğlu et al. (2017) is $Y_i(g,n) = Y_{0i} + \beta_{s(g)} + \gamma \mathbf{1}\{n \geq 1\}$ for $g \geq 1$ and $Y_i(0,n) = Y_{0i}$, with β_s the value added of school s and γ the peer effect.

Example 1 shows why a peer-effects design needs more than one propensity score.

Example 1. *Three native applicants $\{1,2,3\}$ and one migrant applicant m apply for first-grade seats at two schools $\{a,b\}$, each with a single cohort of two seats. No applicant has priority anywhere, ties are broken by lottery, and preferences are*

$$1 : a \succ b \qquad 2 : a \succ b \qquad 3 : b \qquad m : a.$$

Cohort a receives three applications for two seats, and the lottery leaves out each of applicants 1, 2 and m with probability $1/3$. A native who loses at a is seated at b, where there is always room, and m, who ranks only a, is unassigned when she loses. Native applicant 1 is therefore exposed if and only if she and m are both seated at a, which happens when native 2 loses. Her exposure probability is $1/3$, the same as native 2's, while native 3 is never exposed.

Natives 1 and 2 are at school a whenever they are exposed, and when unexposed they are at a with probability $1/2$ (when m loses) and at b with probability $1/2$ (when they lose themselves). A comparison of exposed and unexposed applicants with the same exposure probability, natives 1 and 2, therefore compares school a with an even mixture of a and b: if V_a and V_b are attributes of the two cohorts, such as their value added, the comparison is contaminated by $(V_a - V_b)/2$.

However, notice the problem disappears when exposure is compared within the cohort an applicant is assigned to. Given that native 1 is seated at a, cohort a receives m with probability $1/2$, but given that she is seated at b, her exposure is zero. Thus the difference between her exposure and these conditional probabilities has mean zero whichever school she is assigned to. We later formalize this intuition and show that it generalizes to our main result.

2.3 School Assignment in Chile

Chile's SAE replaced school-run admissions with a single application platform covering every school that receives public funds, municipal and other public schools as well as publicly subsidized private schools, in every grade from pre-kindergarten to the end of secondary school. Only fee-charging private schools and a small delegated-administration sector, which together enroll about a tenth of Chilean students, remain outside. The system was phased in by region: one region in the 2016 process, four more in 2017, all

the others but the Metropolitan Region in 2018, and the Metropolitan Region, which holds almost 40 percent of the country's students, in 2019.[6]

Families rank schools in the second half of the year before entry, and students who continue in their current school do not apply. The Ministry of Education then runs the Student-Proposing Deferred Acceptance algorithm (Gale and Shapley, 1962; Abdulkadiroğlu and Sönmez, 2003) over school-grade cohorts, with priorities set by law: applicants already enrolled at the school come first, then siblings of enrolled students, low-income students up to 15 percent of the grade, children of school staff and former students. Within a priority group, ties are broken by a lottery drawn independently at each school. Families then accept or decline their assignment, a complementary stage places applicants left without a seat, and students may enroll directly in schools with seats remaining.[7] Because the assignment unit is a school-grade cohort, whether a cohort receives a migrant applicant is a lottery outcome, and because the Ministry publishes the application files (rankings, priorities, capacities and lottery numbers), the mechanism can be re-run under alternative lottery draws and its scores computed by simulation, which is what we do. Nationality plays no part in the algorithm, and the admission files do not record it, but we have access to this data via administrative records from the Ministry of Education.

2.4 Data and Descriptive Statistics

A masked national identifier links the admission files to the annual enrollment censuses, the attendance records and the national SIMCE assessments.

Peer type The enrollment censuses have recorded every student's foreign-status since 2008, and we call a student foreign-born if any census records her as foreign. In addition, the enrollment censuses contain an incomplete origin code defining the nationality of the student, which we use to construct a second peer type for applicants explicitly recorded as Venezuelan. We use the fourth-grade SIMCE parent questionnaire, which asks in which country the child was born, to validate both measures. Of the native applicants, 99.3 and 99.0 percent in the two processes have a parent who reports Chile as their country of birth, and of those recorded as Venezuelan, 95.1 and 92.2 percent have a parent who reports Venezuela.

Samples We study the native children who applied for a first-grade seat in the 2018 process, who entered first grade in March 2019, and in the 2019 process, who entered in March 2020.[8] Native applicants total 38,176 and 60,284 in the two processes, and the peer type comprises 722 and 3,008 foreign-born first-grade applicants, of whom 127 and 978 are recorded as Venezuelan. The native applicants who enroll in first grade join cohorts of 64 students on average, taught in classes of about 31 students.

[6] A region joined with its entry grades (pre-kindergarten, kindergarten, first, seventh and ninth grade) and participated with all grades from the following year. The mechanism is described by Correa et al. (2022).

[7] We study the main assignment phase of the mechanism and abstract away from the complementary stage.

[8] The children of the 2019 process moved to remote instruction two weeks after entering school, and most Chilean schools did not resume regular in-person classes until March 2022.

The incumbents of a first-grade cohort are the native students enrolled in the school's kindergarten in the year of the process who did not apply. The lottery does not assign them, and because they are defined from the census taken before it, whether they stay in the school for subsequent years is itself an outcome of interest. There are 13,776 such students in 322 cohorts whose exposure the lottery leaves uncertain in the first process and 39,030 in 894 cohorts in the second.

Outcomes The national SIMCE assessment, which the Agencia de Calidad de la Educación (the Agency) administers, tests math and reading at the end of fourth grade, which the children of the two processes reached in 2022 and 2023, and scores are standardized within each wave over every test taker in the country. The assessment also fields questionnaires answered by the students, with items we group into scales for self-reported victimization by classmates, well-being at school and other attitudes toward schooling.

Descriptive statistics Table 1 describes the data. Column 1 reports means for every native first-grade applicant, and column 2 for the 27,034 applicants, 27 percent of the total, whom the mechanism assigned to a cohort whose exposure the lottery leaves uncertain. These are the applicants who contribute identifying variation.[9] Columns 3 and 4 split them by assigned exposure, and column 5 describes the incumbents. As can be seen in Panel C, exposed applicants are assigned to larger cohorts, with higher foreign-born shares, than unexposed applicants in the same risk set. This is the imbalance Section 3.7 shows the design removes. Panels A and B show exposed and unexposed applicants of similar attendance, age and sex, although the exposed are more likely to have had a foreign-born grademate before the lottery.

3 Exposure Score Theory

Equation (2) describes the randomization of a student to a cohort. We turn to the randomization of cohort composition, which yields another score, and to an instrument that exploits it.

3.1 Different Scores

The probability that a native applicant is exposed combines the probability of her own assignment to each cohort with the probability that her cohort receives a migrant applicant.

Definition 1 (Scores)**.** *For $i \in \mathcal{N}$ and a cohort g with $p_{ig} > 0$, the* conditional cohort score *is $a_{ig} = \mathbb{P}[A_g = 1 \mid D_i(g) = 1]$, the probability that cohort g receives a migrant applicant given that i is assigned there (so $a_{i0} = 0$). The* exposure score *of applicant i is $p_i^Z = \mathbb{P}[Z_i = 1] = \sum_{g \geq 1} p_{ig} a_{ig}$, and the* cohort score *of cohort g is $a_g = \mathbb{P}[A_g = 1]$.*

[9] Note that applicants with degenerate assignment risks (e.g. 100% chance of being assigned to a school) can still contribute identifying variation if the cohort they are assigned to is at risk. This is in contrast to most school value added estimation procedures, where only at-risk students enter the estimation process. The reason behind this difference is that our treatment (the presence of a migrant peer) takes place at the cohort level rather than at the student level.

The conditional cohort score differs from the cohort score because, in a finite market, the event that i is assigned to g is informative about the assignment of the migrant applicants who compete with her for seats there.[10] A condition under which the exposure score and the conditional cohort score are functions of type is the same as the one used by Abdulkadiroğlu et al. (2017) to establish ETE for deferred acceptance. Write $\phi(r)$ for the assignment a mechanism produces from lottery realization r, and $r^{(ij)}$ for the realization in which applicants i and j exchange their lottery numbers. The mechanism is *anonymous* if, whenever $\theta_i = \theta_j$, $\phi_i(r^{(ij)}) = \phi_j(r)$, $\phi_j(r^{(ij)}) = \phi_i(r)$ and $\phi_k(r^{(ij)}) = \phi_k(r)$ for every other applicant k, and the lottery is *symmetric* if r and $r^{(ij)}$ are equally likely whenever $\theta_i = \theta_j$. In other words, an anonymous mechanism will exchange the assignment of two identical applicants if their tiebreakers are exchanged (leaving the rest unchanged), and a symmetric lottery makes the realized tiebreakers and exchanged tiebreakers equally likely for identical applicants.

Proposition 1 (Equal Treatment of Equals for Exposure)**.** *Suppose the mechanism is anonymous and the lottery symmetric, and let $i, j \in \mathcal{N}$ with $\theta_i = \theta_j = \theta$. Then $(D_i(g), N_g)_g$ and $(D_j(g), N_g)_g$ have the same distribution. Consequently, $p_{ig} = p_g(\theta)$, $a_{ig} = \alpha_g(\theta)$ and $p_i^Z = p^Z(\theta)$ are functions of type, and for any attribute W_i fixed under re-randomization, every g, every $a \in \{0, 1\}$ and every w taken by a native applicant of type θ,*

$$\mathbb{P}[D_i(g) = 1,\, A_g = a \mid W_i = w,\, \theta_i = \theta,\, i \in \mathcal{N}] = \mathbb{P}[D_i(g) = 1,\, A_g = a \mid \theta_i = \theta,\, i \in \mathcal{N}]. \tag{4}$$

In particular, $\mathbb{P}[Z_i = 1 \mid W_i = w,\, p^Z(\theta_i) = p,\, i \in \mathcal{N}] = p$.

Because native applicants are not in $\mathcal{A}$, exchanging the lottery numbers of two natives of the same type exchanges their assignments and leaves every migrant applicant's assignment, and hence every N_g, unchanged. The exchange argument fails when one of the two applicants is a migrant, which is why the analysis population is $\mathcal{N}$ and why the peer type must be fixed under re-randomization.

3.2 The Exposure Score Is Not Enough

Conditioning on the exposure score balances every attribute of the applicant, as a propensity score should (Rosenbaum and Rubin, 1983), but not the attributes of the cohort she is assigned to, because exposure is an attribute of that cohort. Holding the exposure score fixed, the applicants who turn out to be exposed are disproportionately those assigned to cohorts with a high conditional cohort score, e.g. cohorts in schools migrant families are more likely to rank. The following proposition quantifies the resulting imbalance.

Proposition 2 (Imbalance under the Exposure Score)**.** *Let $(V_g)_{g \geq 1}$ be cohort attributes fixed under re-randomization, $V_0 = 0$, and $V_{\mu_i} = \sum_g D_i(g) V_g$ the attribute of i's assigned cohort. If $0 < p_i^Z < 1$,*

$$E[V_{\mu_i} \mid Z_i = 1] - E[V_{\mu_i} \mid Z_i = 0] = \sum_{g \geq 1} V_g \, p_{ig} \, \frac{a_{ig} - p_i^Z}{p_i^Z \, (1 - p_i^Z)}. \tag{5}$$

[10] This difference vanishes in a large market (Azevedo and Leshno, 2016).

In Example 1, native 1 reaches cohort a with probability 2/3, where her conditional cohort score is 1/2, and cohort b with probability 1/3, where it is 0, and (5) gives the contamination $(V_a - V_b)/2$. The imbalance does not vanish in large markets, where the conditional cohort scores converge to cohort-specific limits, and pooling applicants of different types averages terms of the form (5), which do not generally cancel out. In the terms of Assumption 1 with the binary mapping, a native applicant's exposure is the pair (μ_i, A_{μ_i}), which takes the value $(g, 1)$ with probability $p_{ig}a_{ig}$ and $(g, 0)$ with probability $p_{ig}(1 - a_{ig})$. The exposure score is the probability of a coarsening of that pair, $Z_i = A_{\mu_i}$, which is why it balances the applicant's attributes but not those of the cohort that comes with her exposure.

Recentering exposure at the exposure score is the recentering of Borusyak and Hull (2023) applied to this design, with the lottery as the shock. Their recentered instrument subtracts from a treatment its expectation over the shocks, here p_i^Z. This also takes the shocks to be as good as randomly assigned with respect to the structural error in the data generating process, so that the shocks move outcomes only through the treatment. However, here the lottery also assigns the applicant's school, which moves her outcome directly. The next proposition gives the covariance of this instrument with the attributes of the applicant and with those of the cohort she is assigned to.

Proposition 3 (Recentering at the Exposure Score)**.** *Let* $\breve{Z}_i = Z_i - p_i^Z$. *For every native applicant* i*:*

(i) $E[\breve{Z}_i] = 0$*, and* $E[\breve{Z}_i W_i] = 0$ *for every attribute* W_i *of the applicant fixed under re-randomization;*

(ii) for cohort attributes $(V_g)_{g\geq 1}$ *fixed under re-randomization, with* $V_0 = 0$*,*

$$E[\breve{Z}_i V_{\mu_i}] = \sum_{g\geq 1} p_{ig}\,(a_{ig} - p_i^Z)\,V_g = \mathrm{Cov}_{p_i}(a_{ig}, V_g), \tag{6}$$

the covariance, over the applicant's assignment distribution $(p_{ig})_{g\geq 0}$ *with* $a_{i0} = 0$*, between the conditional cohort score and the attribute.*

Part (i) is what makes the exposure score a propensity score, as recentering at it, just like conditioning on it, balances everything about the applicant. Part (ii) shows why this single exposure score is not enough. The covariance (6) is the imbalance (5) times $\mathrm{Var}(Z_i) = p_i^Z(1 - p_i^Z)$.[11] It is zero for every cohort attribute if and only if the conditional cohort score equals the exposure score at every cohort the applicant can reach. This condition fails whenever her exposure score is positive and she may be left unassigned, and for an applicant who is always assigned, it requires every cohort she can reach to be equally likely to receive a migrant applicant given that she is assigned there. These conditions are very unlikely to hold in empirical practice.

However, note that both Propositions 2 and 3 highlight a direct connection with the standard formulas for ommited variable bias (OVB). In both cases, the terms are non-zero because of the comovement between a_{ig} and V_g. However, in this case, just as we are able to compute the exposure scores p_i^Z, we can also directly extract the conditional cohort scores a_{ig} from the centralized assignment mechanism. In other words, we can control for the missing variable from which the OVB stems from. This idea is formalized below.

[11] For native 1 of Example 1, (6) is $(V_a - V_b)/9$.

3.3 Recentered Exposure

A solution is to compare exposure within the cohort an applicant is assigned to, by recentering the instrument at the conditional cohort score or, equivalently, by controlling for that score. Given the cohort a native applicant is assigned to, whether that cohort also receives a migrant applicant is a draw from a distribution the mechanism determines, and subtracting its mean from realized exposure leaves a variable whose mean is zero whatever cohort she is assigned to.

Definition 2 (Recentered Exposure)**.** *For $i \in \mathcal{N}$, let $\bar{a}_i = a_{i\mu_i}$ be the conditional cohort score of i's assigned cohort, with $\bar{a}_i = 0$ if $\mu_i = 0$. The* recentered exposure *of i is $\tilde{Z}_i = Z_i - \bar{a}_i = \sum_{g \geq 1} D_i(g)(A_g - a_{ig})$.*

Theorem 1 (Design-Based Orthogonality)**.** *For every native applicant i, $E[\tilde{Z}_i \mid \mu_i] = 0$. Consequently, $E[\tilde{Z}_i \, \psi(\mu_i)] = 0$ for every function ψ of the applicant's own assignment, whose value may also depend on attributes of the applicant or of the cohorts that are fixed under re-randomization. In particular, $\tilde{Z}_i$ is uncorrelated with the applicant's covariates and potential outcomes, with every fixed attribute of her assigned cohort and school, and with every assignment indicator $D_i(g)$.*

Because the conditioning is on the applicant's own assignment, the attributes of her assigned cohort (its size and composition before the lottery, its school's value added) are balanced by design. The theorem holds for any stochastic mechanism, because the conditional cohort scores are defined applicant by applicant, i.e. ETE is needed only to express the scores as functions of type.

3.4 Target Estimand

Theorem 2 (Identified Peer Effect)**.** *Suppose Assumption 1 holds with the binary mapping. Let $\omega_{ig} = p_{ig}\, a_{ig}(1 - a_{ig})$ and suppose $\sum_{i \in \mathcal{N}} \sum_g \omega_{ig} > 0$. Then*

$$\gamma^{\omega} \equiv \frac{E_{\mathcal{N}}[\tilde{Z}_i Y_i]}{E_{\mathcal{N}}[\tilde{Z}_i Z_i]} = \frac{\sum_{i \in \mathcal{N}} \sum_g \omega_{ig}\, \Delta_i(g)}{\sum_{i \in \mathcal{N}} \sum_g \omega_{ig}}, \tag{7}$$

a weighted average of individual peer effects over applicant-cohort pairs, which equals γ under constant effects. Without the binary mapping, the same formula holds with $\Delta_i(g)$ replaced by $E[Y_i(g, N_g) - Y_i(g, 0) \mid D_i(g) = 1, A_g = 1]$.

The weight on a cohort-applicant pair is the probability that i is assigned to g times the conditional variance of g's exposure. Cohorts whose exposure is nearly certain or nearly impossible receive almost no weight, and cohorts that receive a migrant applicant with probability near one half receive the most.

Definition 3 (Effective Number of Randomized Cohorts)**.** *For a set of cohorts with cohort scores a_g, $C_{\text{eff}} = 4 \sum_g a_g(1 - a_g)$.*

We define this measure of the effective number of randomized cohorts because a cohort whose exposure probability is 50% contributes one to C_{eff}, and one whose exposure is degenerate (i.e., always 0 or always 1) contributes nothing. We report this count with our results.

3.5 Incumbents: A Cluster-Randomized Trial

Students already enrolled in a school who do not participate in the match are guaranteed a place in its cohort whatever the lottery does. For such an incumbent, whose cohort $g(i)$ is fixed, assigned exposure is $Z_i = A_{g(i)}$ and the conditional cohort score is the cohort score $a_{g(i)}$. The analog of Assumption 1 is that there are numbers $Y_i(n)$, fixed under re-randomization, with $Y_i = Y_i(N_{g(i)})$. Under the binary mapping, $\Delta_i = Y_i(1) - Y_i(0)$.

Proposition 4 (Incumbents). *For an incumbent i of cohort $g(i)$, let $\tilde{Z}_i = A_{g(i)} - a_{g(i)}$. Then $E[\tilde{Z}_i] = 0$, and $\tilde{Z}_i$ is uncorrelated with every attribute fixed under re-randomization. If $\sum_i a_{g(i)}(1-a_{g(i)}) > 0$, with sums over incumbents, then $\sum_i E[\tilde{Z}_i Y_i] / \sum_i E[\tilde{Z}_i Z_i]$ is the average over incumbents, weighted by $a_{g(i)}(1-a_{g(i)})$, of $E[Y_i \mid A_{g(i)} = 1] - E[Y_i \mid A_{g(i)} = 0]$, which is Δ_i under the binary mapping.*

For incumbents the analysis is a stratified cluster-randomized trial with known, cohort-specific assignment probabilities. Indeed, this analysis is equivalent to the previous one, but setting $p_{ig} = 1$ for $g = g(i)$ and 0 otherwise. We restrict the incumbent sample to cohorts with $0 < a_g < 1$, since other incumbents have $\tilde{Z}_i = 0$.

3.6 Estimating the Scores

We estimate the scores by simulation. For each process we draw $R = 10{,}000$ independent realizations of the lottery, run the mechanism on all applicants for each draw, holding every list, priority and capacity fixed, and record every assignment. Frequencies over draws give $\hat{p}_{ig}$, $\hat{a}_g$, $\hat{a}_{ig}$ and $\hat{p}_i^Z$ for every native applicant and cohort, and hence $\tilde{Z}_i$ at the realized assignment. The same draws deliver our randomization tests.

How well do the simulated scores predict the lottery that took place? Figure 1 compares them with the realized assignment. Panel A groups native applicants into bins of the exposure score and plots the share of each bin whose realized cohort received a migrant applicant against the bin's mean score. Panel B does the same for the conditional cohort score of the realized cohort. In both panels the points lie along the 45-degree line, more tightly for the exposure score, which averages over the cohorts an applicant can reach, than for the cohort score, which is that of a single cohort. Among the 50,435 native applicants whose exposure score is strictly between zero and one, a regression of realized exposure on the score has a slope of 0.977 (standard error 0.016, clustered on the assigned cohort), and, among the 27,034 assigned to a cohort whose conditional cohort score is strictly between zero and one, the slope on that score is 0.942 (0.032). Neither test of a zero intercept and a unit slope rejects, with p-values of 0.321 and 0.194. At the ends of the scale the lottery bears the exposure score out without exception, i.e. the 41,025 native applicants with an exposure score of zero were all unexposed, and the 7,000 with a score of one all exposed.

Table 2 describes the identifying variation. For a foreign-born peer of any origin, the lottery leaves exposure uncertain in 389 of the 3,897 cohorts to which the 2018 process assigned native applicants and in 995 of 5,227 in the 2019 process, and C_{eff} is 195.8 and 483.6. For a peer recorded as Venezuelan it moves exposure in fewer cohorts, with C_{eff} of 81.6 and 290.1.

3.7 Score Balancing Tests

Theorem 1 is exact, but its inputs are simulated, so we validate the quality of the resulting design by analyzing the balance across covariates. We compare uncontrolled differences in the attributes of exposed and unexposed applicants with differences that control for one and for both scores. For each attribute W_i we (i) regress W_i on assigned exposure Z_i with process effects only; (ii) on exposure recentered at the exposure score, $\check{Z}_i = Z_i - p_i^Z$, with a full set of dummies for cells of the exposure score, the comparison of Proposition 3; and (iii) on recentered exposure $\tilde{Z}_i$ with dummies for cells of both scores, the specification of every estimate in the paper. Every attribute W_i is fixed before the lottery, both for the applicant's own attribute, and those of the cohort she is assigned to as it was before the lottery.

Re-estimating the first two regressions on every simulated lottery, with each applicant's assignment and each cohort's exposure as that lottery makes them, gives their average coefficients over the 10,000 draws, i.e. the expected balance. The design is re-estimated on the same lotteries with native assignments held at their realized values, the randomization of our inference procedure (Section 4.1), and each simulated exposure recentered at the cohort score, its mean over the simulated lotteries, which centers the simulated coefficients at zero by construction, as Theorem 1 centers the realized one.

The balance measures reported in Table 3 show large differences in the absence of controls. Exposed applicants are more likely to have had a foreign-born grade-mate before the lottery (0.274 on a base of 0.438), and they are assigned to larger cohorts with more foreign-born students. Controling for the exposure score balances the applicant's own attributes, but not those of the cohort she is assigned to, as Propositions 2 and 3 predict. This can be seen in column 3 of Table 3. The expected differences in the applicant's own attributes, Panels A and B, are zero while exposed applicants are assigned to cohorts with 12.9 more students and a foreign-born share 0.009 higher, in schools 0.082 less likely to have a kindergarten and 0.027 less likely to be rural, differences of 0.28, 0.13, 0.18 and 0.09 standard deviations. Controlling for both scores removes these differences, in expectation by Theorem 1 and at the realized lottery up to the noise of one draw. The last columns of Table 3 report tests for the realized lottery, as would typically be reported for a randomized trial.

Balance of pre-lottery attributes is implied by the mechanism, so these tests check the implementation, whether the simulated scores are accurate enough, rather than the theory. The width of the cells plays no part in that check, as an instrument recentered at the scores themselves is balanced in expectation whatever the width of the cells, which are included for precision, unlike a comparison that conditions on a score only through cells.

4 Using the Score

4.1 Empirical Strategies

We use the recentered exposure to construct instrumental variables estimates of the effect of an immigrant grademate. Section 4.2 turns to classmates. How should these estimates be interpreted? Neither native nor migrant applicants need to enroll where the mechanism assigns them, so assigned exposure moves the exposure a native child

experiences without determining it. Let e_i be the cohort in which native applicant i enrolls and let T_i indicate that a migrant applicant enrolled in e_i, which we call enrolled exposure.

Assumption 2 (Enrollment and Exclusion)**.** *For every $i \in \mathcal{N}$: (a) the applicant's enrollment depends only on her own assignment, $e_i = e_i(\mu_i)$; and (b) there are numbers $Y_i^*(e,t)$, fixed under re-randomization, such that $Y_i = Y_i^*(e_i, T_i)$ for every assignment the mechanism can produce.*

Part (a) says that the school a native enrolls in depends only on her own assignment, i.e. it does not respond to the assignments of migrant applicants, which families do not observe when they accept or decline a seat. Assignment to a cohort that receives a migrant applicant changes the probability that a native enrolls in her assigned school by -0.2 percentage points, with a randomization standard deviation of 1.2. Part (b) is an exclusion restriction extended to exposure: assignments affect outcomes only through the cohort a native enrolls in and the exposure she experiences there. What the lottery does to a native's enrolled exposure at a cohort is summarized by the pair-level compliance $\pi_{ig} = \mathbb{P}[T_i = 1 \mid D_i(g) = 1, A_g = 1] - \mathbb{P}[T_i = 1 \mid D_i(g) = 1, A_g = 0]$, which places no restriction on how migrant applicants take up their seats.

Theorem 3 (Identification under Noncompliance)**.** *Under Assumption 2, if $\sum_i \sum_g \omega_{ig}\pi_{ig} \neq 0$,*

$$\gamma^{IV} \equiv \frac{E_{\mathcal{N}}[\tilde{Z}_i Y_i]}{E_{\mathcal{N}}[\tilde{Z}_i T_i]} = \frac{\sum_{i\in\mathcal{N}} \sum_g \omega_{ig}\pi_{ig}\left[Y_i^*(e_i(g),1) - Y_i^*(e_i(g),0)\right]}{\sum_{i\in\mathcal{N}} \sum_g \omega_{ig}\pi_{ig}}. \tag{8}$$

If $\pi_{ig} \geq 0$ for all pairs, γ^{IV} is a weighted average of the effects of enrolled exposure. Under constant effects, $\gamma^{IV} = \gamma$.

The monotonicity condition $\pi_{ig} \geq 0$, the analog of the condition in Imbens and Angrist (1994), is plausible for a native who enrolls where she is assigned. It can fail for a native who enrolls in another cohort of the same match, because a migrant placed in her assigned cohort is one fewer who can be in the cohort she attends. We directly estimate the first stage for these decliners: it is -0.046 (standard error 0.021) and carries about 3 percent of the absolute first-stage weight. Thus, to the extent that the monotonicity condition fails, it does so for a small fraction of the first-stage weight.

The estimating equations control for both scores through dummies for their cells. Let $d_i^Z(x)$ be dummies for cells of the exposure score, $d_i^a(x)$ dummies for cells of the conditional cohort score of the realized cohort, and X_i the pre-lottery covariates. We estimate

$$T_i = \sum_x \zeta(x)\, d_i^Z(x) + \sum_x \xi(x)\, d_i^a(x) + \delta\, \tilde{Z}_i + X_i'\lambda + \nu_i, \tag{9}$$

$$Y_i = \sum_x \alpha(x)\, d_i^Z(x) + \sum_x \kappa(x)\, d_i^a(x) + \gamma\, T_i + X_i'\eta + \varepsilon_i, \tag{10}$$

by two-stage least squares, with $\tilde{Z}_i$ as the excluded instrument. The cells play the role of strata. They are functions of the applicant's type and assignment, and adding them, or any other function of type, assignment and fixed covariates, leaves the population estimand (8) unchanged. They are included for precision, and they must be coarse enough to leave recentered exposure some variation within them (e.g. dummies for the assigned cohort, the limiting case, leave no variation). We use cells of width 0.02 for

the exposure score and 0.05 for the conditional cohort score, specific to each process. The estimates do not depend on these widths. Indeed, with cells half or twice as wide, or with no cells at all, the pooled reduced form on victimization lies between -0.072 and -0.065. The reduced form, the coefficient on $\tilde{Z}_i$ in a regression of Y_i on $\tilde{Z}_i$ and the controls of (9), is the effect of being assigned a cohort that receives a migrant applicant, and it is the primary estimate in our results table (which we later rescale by the first stage).

As can be seen in Table 2, assignment to a cohort that receives a migrant applicant raises the probability that one enrolls in the native's cohort by 0.316 in the 2018 process and 0.369 in the 2019 process. The first stage is less than one, because a native who declines her seat might not meet an assigned migrant applicant and a migrant applicant who declines hers does not become a grademate, so the 2SLS estimates for applicants are nearly three times the reduced forms. The incumbents' first stages are larger, since an incumbent's enrolled exposure is measured at her school's cohort, so that only migrant applicants' take-up dilutes it.

Inference With a few hundred effectively randomized cohorts, our inference rests on the lottery itself. Under the exposure mappings of Sections 2.2 and 3.5 and the sharp null hypothesis that exposure has no effect on any native, an incumbent's outcome does not depend on the lottery and an applicant's depends on it only through her own assignment, so every outcome is fixed once native assignments are. For each of the 10,000 simulated lotteries we recenter the exposure of each native's realized cohort at its mean over the simulated lotteries, the cohort score a_g, and re-estimate the reduced form with outcomes, covariates and native assignments at their realized values. The randomization p-value is $(1 + b)/(1 + R)$, where b is the number of simulated lotteries whose statistic is at least as large in absolute value as the realized one.

For incumbents this test is exact, up to simulation error in the scores. For applicants it is an approximation, because in a cohort that fills its seats the realized exposure is tied to the native assignments the test holds fixed, i.e. the recentering gives the simulated statistics the center that Theorem 1 gives the realized one.

4.2 Results

Table 4 reports the effect of assignment to a cohort that receives a migrant applicant on six outcomes, for every group we compare. Columns 1 to 3 pool the native applicants of the two processes and then take each process in turn. Columns 4 and 5 divide them by whether the assigned school has a kindergarten, and columns 6 and 7 by whether it teaches first grade in one class or in several. Column 8 reports estimates for the set of incumbents. Panel B reports, for each column, the first stages of a migrant grademate and of a migrant classmate, the native students, the cohorts at risk and C_{eff}. The 2SLS estimates are the reduced form divided by the first stage on the outcome's sample.

Achievement The estimates show no cost of immigrant classmates for native achievement. Pooling the two processes, assignment to a cohort that receives a foreign-born applicant changes the average math and reading score three years later by 0.017 standard deviations, with a randomization standard error of 0.030, and a 95% randomization confidence interval from -0.042 to 0.076. Per enrolled migrant grade-mate, the 2SLS

estimate is 0.046 (standard error 0.081), with an interval from −0.116 to 0.210. The two processes, separate experiments with different applicants and lottery draws, the second adding the Metropolitan Region, yield non-significant effects of 0.012 and 0.021, respectively.

Victimization and well-being Native children report being victimized by their classmates less. The victimization index, which averages the questionnaire's items on how often classmates have mocked, hit or threatened the child and whether she has been afraid of them, falls by 0.070 standard deviations with assignment (randomization standard error 0.024, $p = 0.004$) and by 0.191 per enrolled migrant grade-mate (0.067). The result replicates on both processes, as the estimate is −0.095 in the 2018 process and −0.057 in the 2019 process, each significant on its own randomization distribution. The other indices of the student questionnaire move in the favorable direction or not at all. Self-regulation and conduct rises by 0.052 ($p = 0.074$), and feeling well at school and satisfaction with life rise by 0.036 and 0.023 pooled (non-significant) and by 0.090 and 0.069 in the 2018 process ($p = 0.029$ and 0.085).

Cohorts formed from scratch The lottery does not choose all of a first-grade cohort, as in a school with a kindergarten its children can continue into first grade without applying, so a migrant applicant would join a group whose members have been together for a year or more. In a school without kindergarten, every first grader is new and the match forms the cohort from scratch (94% of the first graders of these schools were applicants of the same process, against 34% where the school has a kindergarten).

Where the match forms the cohort from scratch (column 4), the arrival of a migrant applicant is followed by stronger gains. Native applicants score 0.143 standard deviations higher in math and reading ($p = 0.030$), feel better at school by 0.153 ($p = 0.001$), are more satisfied with life by 0.104 ($p = 0.008$), and report 0.117 less victimization ($p = 0.019$). Where the school has a kindergarten (column 5), victimization falls by 0.064 ($p = 0.017$) and self-regulation rises by 0.070, while achievement, feeling well at school and satisfaction with life do not show significant effects.

Classmates or grade-mates The centralized assignment mechanism assigns applicants to school-grades, and schools then divide them into classes. In a school that teaches first grade in a single class, a migrant applicant who enrolls in the child's grade enrolls in her class by construction, whereas in a school with several classes, she may enroll in another. Panel B of Table 4 shows the consequence for the first stages. In single-class schools the grade-level and class-level first stages are 0.362 and 0.370, nearly equal, while in multi-class schools they are 0.340 and 0.219. If the grade-level assignment affects outcomes only through the child's classmates, Theorem 3 applies with a migrant classmate (instead of grademate) as the treatment and the recentered grade-level instrument unchanged.

Victimization follows the first pattern (columns 6 and 7). The reduced form is −0.109 in single-class schools and −0.059 in multi-class ones, both significant, suggesting these peer effects are mostly driven by interactions with migrant classmates. In addition, when an enrolled migrant is a classmate, achievement shows marginally positive results, as the reduced form is 0.090 ($p = 0.086$).

Children already in the school When studying incumbents, the native children of the school's kindergarten, whose cohort the lottery does not choose but whose composition it changes (column 8), we find that they do not leave. An incumbent whose cohort is assigned a migrant applicant is 2.2 percentage points more likely to be enrolled in the same school a year after entry, on a base of 86.9% ($p = 0.021$), and the 95% randomization interval, from 0.3 to 4.1 points, excludes a fall in retention. The incumbents also describe a more orderly classroom: their index of self-regulation and conduct rises by 0.087 standard deviations ($p = 0.001$). Their scores and their reported victimization do not change significantly, so the decline in victimization is clearest among the children the lottery placed, and the classroom-conduct result among those who were already there.

5 Conclusion

We have shown how to analyze the experiment that a centralized assignment mechanism runs on the composition of its assignment units. Our main theoretical results are three. First, equal treatment of equals extends from assignment to exposure, so the scores that summarize a student's own assignment and the exposure of her cohort are functions of type and can be computed by simulation. Second, conditioning on the probability of exposure, the natural analog of the propensity score, or recentering exposure at it, is not enough, as the comparison it supports balances the student's own attributes but is confounded by those of the cohorts she can reach. This bias does not vanish in a large market. Third, recentering a student's realized exposure at its expectation given her own assignment yields an instrument orthogonal to everything fixed before the lottery and to the quality of her assigned school, which identifies a variance-weighted average of individual peer effects and whose weights deliver an effective number of randomized cohorts (a measure of identifying variation that can be computed before any outcome is examined).

The theoretical results extend beyond immigration, to any fixed attribute of an applicant (e.g. race, disability status, etc.) and to other treatments that are attributes of an assignment unit, and Theorem 1 applies to any stochastic mechanism.

Using our esimtator in the Chilean context, we find no loss in native achievement from immigrant classmates. Assignment to a cohort that receives a migrant applicant does not detectably change the average of math and reading scores three years later or school switching. Native children instead report being victimized by their classmates less. Where the match forms the first-grade cohort from scratch, in schools without a kindergarten, the arrival of a migrant applicant is also followed by higher scores and by native children who also feel better at school, gains that the cohorts growing out of a kindergarten do not show. The children already enrolled in the schools immigrants enter do not leave. If anything, they are more likely to stay.

References

Abdulkadiroğlu, A., Angrist, J. D., Narita, Y., and Pathak, P. A. (2017). Research design meets market design: Using centralized assignment for impact evaluation. *Econometrica*, 85(5):1373–1432.

Abdulkadiroğlu, A., Angrist, J. D., Narita, Y., and Pathak, P. A. (2022). Breaking ties: Regression discontinuity design meets market design. *Econometrica*, 90(1):117–151.

Abdulkadiroğlu, A., Angrist, J. D., and Pathak, P. A. (2014). The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica*, 82(1):137–196.

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

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30:98–108.

Azevedo, E. M. and Leshno, J. D. (2016). A supply and demand framework for two-sided matching markets. *Journal of Political Economy*, 124(5):1235–1268.

Borusyak, K. and Hull, P. (2023). Nonrandom exposure to exogenous shocks. *Econometrica*, 91(6):2155–2185.

Borusyak, K., Hull, P., and Jaravel, X. (2022). Quasi-experimental shift-share research designs. *Review of Economic Studies*, 89(1):181–213.

Carrell, S. E., Sacerdote, B. I., and West, J. E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica*, 81(3):855–882.

Correa, J., Epstein, N., Epstein, R., Escobar, J., Rios, I., Aramayo, N., Bahamondes, B., Bonet, C., Castillo, M., Cristi, A., Epstein, B., and Subiabre, F. (2022). School choice in Chile. *Operations Research*, 70(2):1066–1087.

Epple, D. and Romano, R. E. (2011). Peer effects in education: A survey of the theory and evidence. In Benhabib, J., Bisin, A., and Jackson, M. O., editors, *Handbook of Social Economics*, volume 1B, pages 1053–1163. North-Holland, Amsterdam.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly*, 69(1):9–15.

Hoxby, C. M. (2000). Peer effects in the classroom: Learning from gender and race variation. Working Paper 7867, National Bureau of Economic Research.

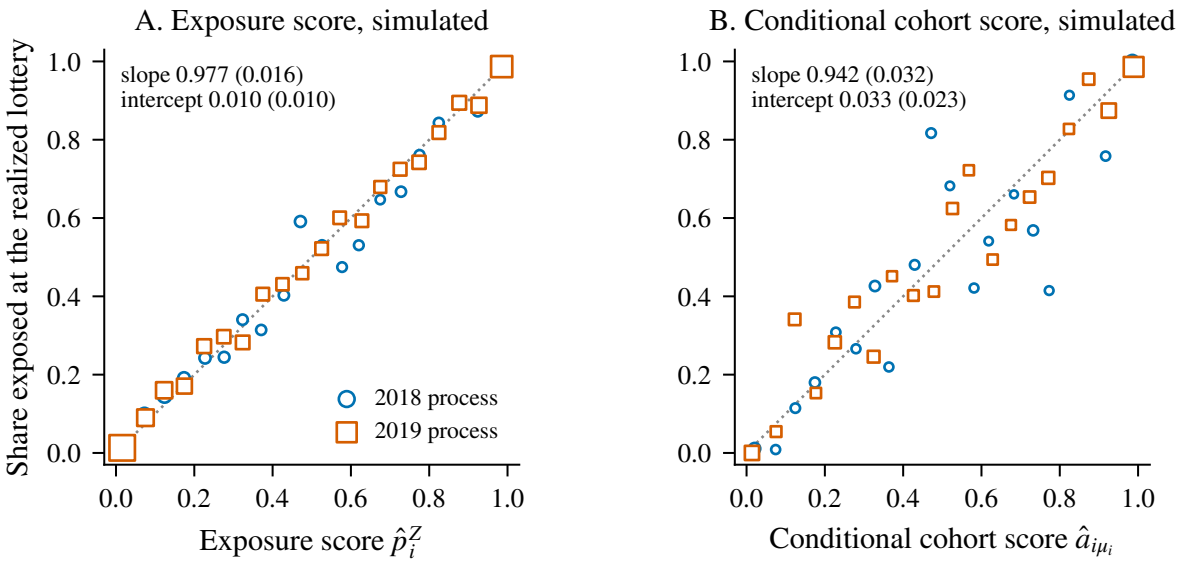
Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3):531–542.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics*, 116(2):681–704.

Figure 1: Simulated scores and realized lottery shares



Note: This figure shows the simulated scores and the realized lottery shares. Each marker is a bin of width 0.05 of the score on the horizontal axis, for the native first-grade applicants of the 2018 process or the 2019 process whose score is strictly between zero and one. It plots the share of the bin's applicants whose realized cohort received a foreign-born applicant against the bin's mean score, with marker area increasing in the number of applicants in the bin. Panels A and B use the simulated exposure score over all native applicants and the conditional cohort score of the realized cohort over those the mechanism assigned. The dotted line is the 45-degree line, and each panel reports the intercept and slope of a regression of realized exposure on the score over the applicants it plots, both processes pooled, with standard errors clustered on the assigned cohort in parentheses.

Table 1: Descriptive Statistics

	Native applicants				
		Assigned a cohort at risk			
	All	All	Exposed	Not exposed	Incumbents
	(1)	(2)	(3)	(4)	(5)
A. Application					
Schools listed	3.31	3.69	3.71	3.65	n.a.
Lists her current school	0.304	0.344	0.327	0.366	n.a.
Sibling priority	0.283	0.340	0.324	0.359	n.a.
Low-income priority	0.492	0.425	0.429	0.421	n.a.
B. Before the lottery					
Female	0.457	0.464	0.468	0.460	0.507
Age relative to grade	−0.80	−0.79	−0.78	−0.79	−0.79
Enrolled in school	0.901	0.918	0.922	0.912	1.000
In a public school	0.260	0.250	0.262	0.236	0.250
Attendance rate	0.902	0.899	0.898	0.900	0.880
Chronically absent	0.087	0.097	0.102	0.090	0.149
Foreign-born share, her cohort	0.027	0.033	0.038	0.027	0.030
Foreign-born grade-mate	0.529	0.647	0.690	0.594	0.599
Her school had no first grade	0.608	0.587	0.601	0.571	0.000
C. The assigned cohort					
Enrollment	66.6	84.1	88.1	79.3	93.0
Foreign-born share	0.030	0.026	0.033	0.018	0.029
Venezuelan share	0.006	0.005	0.007	0.004	0.005
Public school	0.368	0.234	0.271	0.190	0.250
Rural school	0.105	0.019	0.019	0.019	0.027
School's fourth-grade score, 2018	−0.07	0.05	0.01	0.10	0.08
School charges a copayment	0.270	0.433	0.371	0.509	0.368
School's socioeconomic group high	0.159	0.271	0.225	0.325	0.229
School has a kindergarten	0.713	0.684	0.660	0.714	1.000
Students	98,460	27,034	14,761	12,273	52,806

Note: This table reports means for the native children who applied for a first-grade seat in the 2018 and 2019 processes (columns 1 to 4) and for the incumbents who did not apply (column 5). Column 2 is the applicants assigned to a cohort at risk, whose probability of receiving a foreign-born applicant, given that they are assigned there, is strictly between zero and one. Columns 3 and 4 split them by assigned exposure. Panel A describes the application, which incumbents did not file (n.a.). Panel B describes the child in the year of the lottery (attendance averaged over that year and the year before, chronic absence in both, the foreign-born share of her school-grade in 2018, and, in that year's census, whether her school-grade held a foreign-born child and whether her kindergarten or special-education preschool had no first grade). Panel C describes the cohort she is assigned to, for an incumbent the first grade of her own school, as it was before the lottery: its enrollment and composition in 2018, and its school, whose sector, rurality and kindergarten are those of the census of the year of the lottery, fourth-grade score (the mean in national standard deviations of its fourth graders' scores) and socioeconomic group of the 2018 national assessment, and copayment status. Every attribute in panels B and C is measured alike for applicants and incumbents. The two processes are pooled.

Table 2: Identifying Variation

Peer type	Process	Students (1)	Cohorts at risk (2)	C_{eff} (3)	Assigned exposure (4)	First stage (5)	Clustered F (6)
A. Native applicants							
Foreign-born, any origin	2018	38,176	389	195.8	0.247	0.316 (0.068)	27.0
	2019	60,284	995	483.6	0.376	0.369 (0.047)	78.4
Venezuelan	2018	38,176	177	81.6	0.089	0.383 (0.100)	22.4
	2019	60,284	610	290.1	0.172	0.395 (0.064)	54.3
B. Incumbents							
Foreign-born, any origin	2018	13,776	322	152.5	0.381	0.442 (0.084)	28.0
	2019	39,030	894	425.7	0.427	0.455 (0.059)	74.2
Venezuelan	2018	6,563	151	67.5	0.300	0.376 (0.129)	8.5
	2019	22,357	548	250.6	0.348	0.419 (0.073)	32.3

This table describes the variation the lotteries generate in each process. Cohorts at risk are the first-grade cohorts to which the mechanism assigned native applicants (Panel A) or which hold incumbents (Panel B) and whose cohort score a_g is strictly between zero and one. $C_{\text{eff}} = 4\sum_g a_g(1-a_g)$ is the effective number of randomized cohorts (Definition 3). Assigned exposure is the share of assigned applicants, or of incumbents, whose cohort receives a migrant applicant of the peer type. The first stage is the coefficient of enrolled exposure on recentered exposure with the controls of equation (9). Randomization standard errors are in parentheses, and F is the conventional first-stage F-statistic, from a standard error clustered on the assigned cohort.

Table 3: Covariate Balance

	Mean, unexposed	Expected difference			Realized difference		
		No controls	Exposure score	Both scores	No controls	Exposure score	Both scores
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Application							
Schools listed	3.17	0.50	0.00	0.00	0.50***	0.02	−0.02
Lists her current school	0.300	0.010	0.000	0.000	0.013***	0.012	0.010
Sibling priority	0.271	0.041	0.000	0.000	0.038***	−0.005	−0.026*
Low-income priority	0.508	−0.050	0.000	0.000	−0.049***	0.003	0.007
B. Before the lottery							
Female	0.450	0.023	0.000	0.000	0.024***	0.003	0.003
Age relative to grade	−0.80	0.01	0.00	0.00	0.01***	0.00	0.00
Enrolled in school	0.888	0.037	0.000	0.000	0.043***	0.019*	0.026
In a public school	0.245	0.055	0.000	0.000	0.055***	0.002	−0.004
Attendance rate	0.904	−0.009	0.000	0.000	−0.008***	0.005**	0.006*
Chronically absent	0.080	0.024	0.000	0.000	0.024***	0.001	0.004
Foreign-born share, her cohort	0.018	0.030	0.000	0.000	0.030***	0.000	−0.001
Foreign-born grade-mate	0.438	0.274	0.000	0.000	0.269***	−0.007	−0.009
Her school had no first grade	0.605	0.018	0.000	0.000	0.015***	−0.011	−0.007
C. The assigned cohort							
Enrollment	55.3	31.9	12.9	0.0	31.9***	12.8***	−0.9
Foreign-born share	0.014	0.047	0.009	0.000	0.047***	0.009***	0.000
Venezuelan share	0.003	0.008	0.003	0.000	0.008***	0.003***	0.000
Public school	0.363	0.024	−0.003	0.000	0.026***	0.009	0.028
Rural school	0.146	−0.118	−0.027	0.000	−0.117***	−0.024***	0.009
School's fourth-grade score, 2018	−0.08	0.04	0.03	0.00	0.03***	0.01	−0.04
School charges a copayment	0.270	0.011	0.026	0.000	−0.003	−0.023	−0.088*
School's socioeconomic group high	0.155	0.021	0.026	0.000	0.006	−0.022	−0.081
School has a kindergarten	0.749	−0.124	−0.082	0.000	−0.113***	−0.043	0.062
Joint test p, both processes					<0.001	<0.001	0.845

This table reports differences in pre-lottery attributes between native applicants whose assigned cohort does and does not receive a foreign-born applicant, the two processes pooled. Columns 2 and 5 regress the attribute on assigned exposure Z_i with process effects. Columns 3 and 6 on exposure recentered at the exposure score, $Z_i - p_i^Z$, with dummies for cells of the exposure score. Columns 4 and 7, the design, on recentered exposure $\tilde{Z}_i$ with dummies for cells of both scores. Columns 5 to 7 are the realized coefficients. Columns 2 and 3 average the coefficient over the 10,000 simulated lotteries, each redrawing every assignment. Column 4 averages it over the same lotteries with native assignments held at their realized values and exposure recentered at the cohort score. Stars compare each realized coefficient, against the normal, with the standard deviation of the simulated coefficients its expected column averages: *significant at 10%; **significant at 5%; ***significant at 1%. The last row gives the joint p-value (columns 5 to 7).

Table 4: Effects of Assignment to a Cohort That Receives a Migrant Applicant

	Native applicants							
	Processes			Kindergarten		First-grade classes		
	Both	2018	2019	No	Yes	One	Two or more	Incumbents
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Reduced form: assignment to a cohort that receives a migrant applicant								
Reading and mathematics	0.017	0.012	0.021	0.143**	−0.034	0.090*	−0.015	−0.049
	(0.030)	(0.047)	(0.038)	(0.066)	(0.030)	(0.053)	(0.036)	(0.035)
Victimization by classmates	−0.070***	−0.095**	−0.057**	−0.117**	−0.064**	−0.109**	−0.059**	−0.020
	(0.024)	(0.046)	(0.028)	(0.050)	(0.027)	(0.048)	(0.029)	(0.021)
Feeling well at school	0.036	0.090**	0.014	0.153***	0.008	0.088	0.019	0.032
	(0.025)	(0.042)	(0.031)	(0.048)	(0.029)	(0.055)	(0.029)	(0.022)
Satisfaction with life	0.023	0.069*	0.003	0.104***	0.003	0.045	0.010	0.024
	(0.022)	(0.040)	(0.027)	(0.040)	(0.028)	(0.047)	(0.027)	(0.019)
Self-regulation and conduct	0.052*	0.039	0.061	0.082	0.070**	0.022	0.063*	0.087***
	(0.029)	(0.042)	(0.037)	(0.055)	(0.034)	(0.058)	(0.036)	(0.027)
Same school, $t+1$	−0.008	−0.034	0.004	−0.018	0.007	0.017	−0.010	0.022**
	(0.011)	(0.027)	(0.009)	(0.024)	(0.010)	(0.016)	(0.013)	(0.010)
B. First stages and design								
First stage, migrant grade-mate	0.351	0.316	0.369	0.457	0.295	0.362	0.340	0.452
	(0.039)	(0.068)	(0.047)	(0.092)	(0.039)	(0.062)	(0.047)	(0.049)
First stage, migrant classmate	0.254	0.211	0.274	0.321	0.227	0.370	0.219	
	(0.028)	(0.047)	(0.034)	(0.072)	(0.028)	(0.062)	(0.029)	
Native students	98,460	38,176	60,284	25,053	62,365	33,524	53,662	52,806
Cohorts at risk	1,384	389	995	179	1,205	575	807	1,216
C_{eff}	679.3	195.8	483.6	105.9	573.4	286.1	392.3	578.2

This table reports the effect of assignment to a first-grade cohort that receives a foreign-born applicant: the coefficient on recentered exposure $\tilde{Z}_i$ in a regression with dummies for cells of the exposure score and of the conditional cohort score, the pre-lottery covariates and, where processes are pooled, process effects, the cells being specific to each process. Columns 1 to 7 are native applicants: both processes and each; those assigned to a school without a kindergarten, whose first-grade cohort the match forms from scratch, and with one; and those assigned to a school that teaches first grade in one class and in two or more, in the entry year. Column 8 is the set of incumbents, whose instrument is their cohort's recentered exposure $A_{g(i)} - a_{g(i)}$, with cells of the cohort score and controls. The questionnaire indices and the average math and reading score of the fourth-grade assessment are in standard deviations. Same school is enrollment a year after entry in the school of the entry year, among the applicants enrolled that year, and in the school of the kindergarten for incumbents. The first stage is the coefficient of an indicator that a migrant applicant enrolled in the child's grade or in her own class, on the whole sample of the column. The 2SLS estimate divides the reduced form by the first stage on the outcome's sample. Randomization standard errors in parentheses. Stars compare each estimate with its standard error against the normal distribution: *significant at 10%; **significant at 5%; ***significant at 1%. Cohorts at risk are those whose probability of receiving a migrant applicant is strictly between zero and one, and $C_{\text{eff}} = 4\sum_g a_g(1-a_g)$ over them.