Disentangling School and Peer Effects: A Non-linear Approach with School Lotteries *

Mariana Laverde Boston College
Minseon Park University of Michigan
Suk Joon Son University of Tokyo

September 2026

Abstract

We estimate peer effects and peer-adjusted school effects in a model in which student outcomes depend on the full distribution of peer achievement. Building on school choice lottery-based methods for identifying school effects, we show how centralized assignment lotteries can be used to jointly identify peer-adjusted school effects and peer effects by exploiting random variation in both school assignment and peer composition. Functional principal component analysis provides a tractable low-dimensional representation of peer-achievement distributions. Applying the framework to New York City middle schools, we find that shifting the observed peer distribution 0.5 standard deviations toward higher-achieving peers lowers predicted sixth-grade achievement by 0.029 standard deviations on average, with larger effects for students with above-median baseline achievement. Longer-run outcomes related to graduation and college enrollment show a different pattern, with suggestive evidence of positive effects from higher-achieving peers, pointing to potentially different channels of peer influence across outcomes. Accounting for peer composition changes estimated school value-added relatively little: peer-adjusted and conventional school effects have a correlation of 0.93. We then study optimal student assignment when achievement depends on both peer effects and peer-adjusted school effects. The best feasible reassignment raises predicted achievement by about 0.009 standard deviations; about 70 percent of the gain comes from improved student–school matches and the remainder from changes in peer composition.

Keywords: Peer effects; school value added; functional principal component analysis; optimal school assignment.

JEL Codes: I21, C31, C14, D61, I28

*This paper benefited from discussions with Nikhil Agarwal, Francesco Agostinelli, Jesse Bruhn, Yoosoon Chang, Angela Crema, Bo Hu, Peter Hull, Hidehiko Ichimura, Daiji Kawaguchi, Ryo Okui, Parag Pathak, Chris Taber, Chris Walters, and seminar participants at the UM-MSU-UWO Labo(u)r Day Conference, Re-imagining Educational Measurement at the Annenberg Institute, Toulouse School of Economics 1st Human Capital Conference, Barcelona School of Economics Summer Forum Workshop on Networks, Tokyo Labor Workshop, Korea Economic Association Workshop at Waseda, SOLE, and ERATO Meeting. Emails: laverdem@bc.com, minseonp@umich.edu, and sjson@e.u-tokyo.ac.jp.

1. Introduction

Designing education policy requires understanding how schools and peers separately contribute to student learning. Many policies—including affirmative action in admissions, accountability-driven school closures, and expansions of school choice through charter schools or vouchers—simultaneously change the schools students attend and the peers they encounter. Separating these channels is therefore necessary both to evaluate the effects of such policies and to understand which margin offers greater scope for improving achievement. If peer composition is a primary driver of achievement, policies that change who learns with whom—through affirmative action or tracking, for example—may generate substantial gains. If instead peer-adjusted school effects account for most of the relevant variation, policies that redirect students toward more effective schools—such as accountability policies that close low-performing schools—or better match students to schools that are particularly effective for them may offer greater scope for improving achievement.

Despite this policy relevance, evidence that separately identifies the contributions of schools and peers remains limited. Estimating these separate contributions poses two challenges. First, identifying causal peer and peer-adjusted school effects requires variation in both school assignment and peer composition that is unrelated to students' unobserved ability. In addition, peer composition must also vary independently of unobserved school quality; otherwise, its estimated effect conflates the influence of classmates with the quality of the school they share. The second challenge is dimensionality. Estimating separate effects for many schools requires recovering a large set of school-specific parameters. Allowing outcomes to depend on the full distribution of peer achievement also requires representing that distribution in a way that is tractable for estimation while preserving potentially relevant features of the distribution. The linear-in-means model offers a parsimonious solution by reducing the distribution to its mean, but doing so may miss relevant dimensions of peer composition and, importantly, mechanically renders the aggregate contribution of peers to student outcomes invariant to how students are grouped.

In this paper, we develop and estimate a model of school and peer effects that allows student outcomes to depend on the full distribution of peer achievement while addressing both challenges. To address identification, we exploit two sources of lottery-induced variation in centralized school-assignment systems. A student's own lottery draw can generate variation in the

school she is assigned to (Angrist et al., 2017), while other students' draws generate variation in the composition of students assigned to the same school that can be isolated from systematic differences in school effectiveness. To address dimensionality, we use functional principal component analysis (FPCA) (Chang et al., 2023) to represent peer-achievement distributions with a small number of components, allowing us to estimate peer effects that depend flexibly on the full peer-achievement distribution, jointly with individual school effects.

We apply this framework to the centralized middle-school assignment in New York City, the largest public school district in the United States, to estimate the effect of peers and the separate contribution of schools to learning. We use these estimates for two purposes. First, we assess the extent to which conventional measures of school value-added reflect the peer composition associated with each school rather than school-specific inputs and practices. Second, we quantify the learning gains attainable by optimizing student assignments across schools, where each assignment affects learning through both the destination school's peer-adjusted effect and the peer composition it generates. Our estimates show higher-achieving peers need not improve achievement and point to a smaller role for peer composition than for peer-adjusted school effects in shaping student outcomes. Consistent with this, the gains from optimal reassignment operate primarily through improved student–school matches rather than peer effects.

We model student achievement as the sum of a linear functional of the peer-achievement distribution and a peer-adjusted school effect. We measure peers' prior achievement relative to the focal student's own prior achievement, so that peer effects depend on the distance between a peer's achievement and the student's baseline achievement. A peer-effect function describes how predicted achievement changes as mass shifts across different points of the relative-peer-achievement distribution. It captures both the effects of direct interactions among students and changes in the school inputs a student experiences as peer composition changes, such as teachers adapting instruction to a cohort's preparation or changes in teacher-student assignments (Lazear, 2001; Sacerdote, 2011). The peer-adjusted school effect captures learning impacts that are not explained by lottery-induced changes in peer composition, including the effects of school leadership, teacher quality, and established pedagogical models.

Our identification strategy leverages two distinct sources of lottery-induced variation: variation

in school assignment and variation in peer composition that is independent of peer-adjusted school effectiveness. Together, these sources of variation allow us to separately identify the peer-effect function and peer-adjusted school effects. To implement this strategy, we simulate the mechanism to construct each student's school-assignment probabilities and expected co-assigned peer composition, and condition on these objects together with student characteristics. Because lotteries randomize assignment rather than enrollment, our OLS implementation requires that, conditional on these controls, the schools students enroll in and the peers they encounter be unrelated to unobserved determinants of achievement. Building on Borusyak and Hull (2023) and Angrist et al. (2024), we formalize this identification argument and state the conditional-independence conditions required for our OLS implementation. The lottery-induced school-assignment and peer-composition shocks excluded from the estimating equation then provide additional moment restrictions that we use to test these identifying assumptions.

To incorporate the full peer-achievement distribution into an estimable achievement equation, we use FPCA to represent each enrolled-peer density as a linear combination of a small number of common basis functions, with student-specific loadings (Chang et al., 2023). Among representations of a given dimension, FPC captures the largest possible share of across-student variation in peer composition, allowing us to preserve as much variation in peer composition as possible while maintaining tractability. We retain three components for enrolled-peer densities, which explain 95 percent of their observed variation. Our estimating equation then relates sixth-grade achievement to enrolled-school indicators and co-enrolled peer composition FPCA loadings, controlling for baseline achievement, demographics, school-assignment probabilities, and FPCA loadings for expected co-assigned peer composition. The coefficients on the enrolled-peer loadings map back into the peer-effect function, while the school coefficients identify peer-adjusted school effects.

Our estimates reveal that higher-achieving peers do not uniformly raise achievement. In the full sample, the peer-effect function is broadly downward sloping, implying that exposure to higher-achieving peers can reduce achievement. A uniform 0.5-standard-deviation shift of the observed peer distribution toward higher-achieving peers lowers average predicted sixth-grade achievement by 0.029 standard deviations, while a shift of the same size toward lower-achieving peers raises predicted achievement by 0.031 standard deviations. Under the corresponding linear-

in-means specification, the same 0.5-standard-deviation shift toward higher-achieving peers lowers predicted achievement by 0.024 standard deviations. The flexible specification reveals modest departures from the constant-slope restriction while preserving a broadly monotone relationship. Full-sample estimates mask relevant heterogeneity. For high-achieving students, a 0.5-standard-deviation rightward shift lowers predicted achievement by 0.055 standard deviations, compared with 0.011 standard deviations for low-achieving students.

Accounting for peer composition changes estimated school effects relatively little. In the homogeneous model, the estimated standard deviation of peer-adjusted school effects is 0.162 student-achievement standard deviations, compared with 0.154 when peers are omitted. School-specific estimates from the two models have a correlation of 0.93. The same pattern holds when school effects vary with baseline achievement, although differences are largest for high-achieving students. For these students, peer-adjusted school effects are more dispersed than school effects estimated without accounting for peers. More effective schools tend to enroll higher-achieving peers, whose estimated contribution to achievement is negative for these students. When peer composition is omitted, this negative contribution is absorbed into the estimated school effects, slightly compressing estimated dispersion. Even so, accounting for peers does not substantially reorder schools in terms of their value-added.

Taking our estimates of peer and school effects, we consider the optimal reassignment of students across schools to maximize achievement while preserving school capacities. To limit extrapolation beyond observed peer environments, we restrict each school's counterfactual composition to a mixture of observed school compositions. The best feasible allocation raises predicted achievement by about 0.009 student-achievement standard deviations. About 70 percent of this gain comes from matching achievement groups to schools with larger peer-adjusted effects for their group, and the remaining 30 percent comes through the peer channel. The scope for improving achievement through peer composition is smaller than through student–school matching. Across feasible assignments, the difference between the allocations that maximize and minimize the peer contribution is about 0.028 standard deviations, compared with about 0.043 standard deviations for the school-match contribution. On each margin, the observed assignment already attains roughly two-thirds of the range between the worst and best feasible allocations. To make optimization

tractable, we reformulate the planner's problem over school-level baseline-achievement distributions, rather than over a discrete student-level assignment, and parameterize each distribution by a small number of loadings on the leading FPCs estimated from observed school-level densities. We restrict each school's counterfactual composition to a mixture of observed school compositions, limiting extrapolation beyond observed peer environments. The resulting continuous, low-dimensional problem is amenable to standard nonlinear optimization.

This paper contributes to the literature on peer effects by providing a tractable framework to estimate a model in which student outcomes depend flexibly on the full distribution of peer achievement, rather than on a prespecified summary of that distribution, such as its mean. Our framework nests the canonical linear-in-means model. To estimate the flexible peer effect function, we use FPCA to obtain a tractable representation. Existing work has departed from the linear-in-means framework by allowing outcomes to depend on selected moments of the peer distribution, on the shares of peers in particular achievement groups, on interactions between these measures and the student's own achievement, or on nonlinear parametric aggregations of peer outcomes (Hoxby and Weingarth Salyer, 2006; Lavy et al., 2012; Imberman et al., 2012; Burke and Sass, 2013; Boucher et al., 2024; Tincani, 2026). Relative to these approaches, our estimation uses a data-driven representation that captures multiple dimensions of peer composition rather than selecting particular features of the distribution ex ante.[1]

We contribute to a literature that seeks to understand which inputs, practices, and mechanisms account for variation in school effectiveness. We focus on the role of peers. Existing work has separated peer composition from other school effects in several ways. Jackson (2013) decomposes the effect of attending a more selective school into a peer component and a residual school component, while Bastos et al. (2022) isolate peer-adjusted school effects in a setting where universal random assignment balances peer composition across schools. Other work separately estimates peer and school effects using instrumental-variable or student-mobility designs (Allende, 2019; Woodcock et al., 2025). These approaches generally summarize peer composition using a small number of prespecified characteristics. Our paper adds to this literature by jointly estimating school-specific peer-adjusted effects and a flexible peer-effect function, allowing us to assess how much measured

[1]The baseline specification is not fully unrestricted; in particular, it does not generally nest peer effects that depend nonlinearly on moments of the peer distribution, such as the variance.

school effectiveness reflects peer composition rather than school-specific inputs and practices.

Methodologically, our paper builds on a literature that establishes conditions and methods for identifying school effects using centralized assignment lotteries (Abdulkadiroğlu et al., 2017; Angrist et al., 2017, 2024). We extend this logic by recognizing that centralized assignment lotteries generate random variation not only in the school a student is assigned to, but also in the peers she is assigned alongside. Bibler et al. (2026) exploit one margin of this peer-assignment variation to estimate spillovers on students left behind when lottery winners exit their assigned schools. We formalize conditions under which school-assignment and peer-composition variation can be used jointly to identify peer-adjusted school effects and peer effects, showing when conditioning on school-assignment probabilities and expected co-assigned peer composition isolates the relevant lottery-induced variation.

Finally, we connect our estimates to the design of optimal student assignments. Abdulkadiroglu et al. (2026) study optimal reassignment when gains arise from rich student–school match effects. We instead allow assignments to affect achievement through heterogeneous peer-adjusted school effects and peer effects that depend flexibly on endogenous peer composition. This makes the problem of assigning students across schools generally intractable: the value of assigning one student depends on who else is assigned to the same school, and the flexible peer-effect structure need not admit a simple sorting rule. We show how our low-dimensional FPCA representation can make this problem tractable by reformulating the planner's problem over school-level achievement distributions, parameterized by their FPCA loadings, rather than individual student assignments. This transformation does not generally make the problem convex, but it makes it amenable to standard nonlinear optimization tools. Other work has studied assignment across peer groups under more restrictive representations of peer effects, including discrete peer types or shares (Carrell et al., 2013; Graham et al., 2025) and selected moments of the peer distribution to evaluate reassignment rules rather than solve the unrestricted optimum (Booij et al., 2017).

The rest of the paper is organized as follows. Section 2 presents the model of school and peer effects. Sections 3 and 4 develop the identification strategy and FPCA approach. Section 5 describes the setting, data, estimation, and validation tests. Section 6 presents the estimates, Section 7 studies the optimal assignment of students to schools, and Section 8 concludes.

2. A Model of Peer and School Effects

We model student achievement as the sum of a school effect and a peer effect that depends flexibly on the distribution of peer quality. Let $f_i(r)$ denote the distribution of peer quality among student i's peers, where r indexes its possible values. We construct this measure from lagged test scores, as defined below. We assume that the achievement of student i in year t, when she attends school j, is given by

$$y_{it} = \alpha + \mu_{jt} + \int \beta(r)\, f_i(r)\, dr + \varepsilon_{it}. \tag{1}$$

Here, α is the average potential achievement in the district, μ_{jt} is a school effect that may, but need not, vary over time, and ε_{it} captures idiosyncratic determinants of student achievement, including student ability. The function $\beta(r)$ describes the contribution to student achievement per unit mass of peers at peer-quality value r. Thus, the term $\int \beta(r) f_i(r)\, dr$ aggregates these contributions over the full distribution of peer quality. We interpret μ_{jt} as a peer-adjusted school effect: it captures the contribution of attending school j in year t to student achievement after accounting for the contribution of peers.

We measure peer quality relative to the focal student's own prior achievement. Specifically, peer k's relative-achievement value for student i is the difference between peer k's lagged achievement and student i's lagged achievement, $r = y_{k,t-1} - y_{i,t-1}$. Thus, $f_i(r)$ describes the distribution of differences between each peer's lagged test score and student i's own lagged test score. Positive values of r correspond to peers with higher prior achievement than student i, while negative values correspond to peers with lower prior achievement. This choice reflects the idea that the effects of a given peer may depend on the focal student's own academic preparation. For example, opportunities to learn from classmates and the suitability of classroom instruction may depend on how far ahead or behind those classmates are. Under this specification, peer effects depend on the direction and the magnitude of the achievement difference, rather than assuming that a peer with a given absolute level of prior achievement has the same effect on every student.

Under either specification, the peer effect function $\beta(r)$ admits a counterfactual interpretation.[2]

[2] Appendix C provides a formal derivation.

Holding the school and individual components of student i's achievement fixed, consider reallocating mass within the peer-achievement distribution. For any two points r and r' in the support of f_i and any mass m, the expression $\big[\beta(r') - \beta(r)\big]m$ gives the change in student i's predicted test score induced by replacing a mass m of peers at r with an equal mass of peers at r'. This interpretation also highlights that only differences in $\beta(r)$ are identified: because $f_i(r)$ integrates to one, any constant shift in the peer effect function is absorbed by the intercept α, so its level is not separately identified. We therefore normalize $\beta(0) = 0$, so that $\beta(r)$ is interpreted relative to the effect of peers at $r = 0$.

Our model nests the widely used linear-in-means model as a special case. In the linear-in-means model, peer effects depend only on the average prior achievement of a student's peers, rather than on other features of the peer-achievement distribution. The linear-in-means model is obtained by imposing that the peer effect function is linear in peer quality: $\beta(r) = br$, where b is a scalar. Equation (1) then becomes

$$y_{it} = \mu_{jt} + b\bar{r}_i + \varepsilon_{it}, \quad \text{where} \quad \bar{r}_i = \int r\, f_i(r)\, dr.$$

When peer quality is measured by peers' absolute lagged achievement, $\bar{r}_i$ represents the average lagged achievement of i's peers, and this specification reduces to the canonical linear-in-means model. Our baseline specification instead measures peer quality relative to student i's own lagged achievement. Under this definition, the corresponding linear-in-means specification is equivalent to the canonical model based on average absolute peer achievement, provided that student i's own lagged achievement enters the outcome equation.[3]

More generally, our framework accommodates peer-effect specifications that are linear combinations of raw moments of the peer-quality distribution. By setting the peer effects function to be a linear combination of powers of r, $\beta(r) = \sum_{l=1}^{L} b_l r^l$, the peer component of the model can then be written as a linear combination of the first L raw moments of f_i:

[3]Suppressing time subscripts for simplicity, let $\bar{y}_{-i}$ denote the average lagged test score of student i's peers. Because relative peer achievement is $r = y_k - y_i$, mean relative peer achievement satisfies $\bar{r}_i = \bar{y}_{-i} - y_i$. Hence, $b\bar{y}_{-i} = b\bar{r}_i + by_i$. The conventional peer term based on average absolute achievement therefore equals the relative-achievement peer term plus a term proportional to student i's own lagged achievement.

$$\int \beta(r) f_i(r)\, dr = \sum_{l=1}^{L} b_l \int r^l f_i(r)\, dr.$$

The baseline specification does not, however, generally nest effects that depend nonlinearly on raw moments of the peer-quality distribution. For example, it does not nest a specification with an effect of peer-quality variance.[4] When theory or prior evidence suggests that such features matter, they can be included as additional regressors alongside the functional peer-effect term.

Having described the peer component of the model, we next clarify how we conceptually distinguish peer effects from peer-adjusted school effects. We define peer effects to include effects arising both from direct interactions among students and from endogenous school responses to random year-to-year changes in peer composition. For example, teachers may adapt instruction to students' academic preparation. Because these responses are induced by cohort composition, we attribute them to peers, consistent with the treatment of peer effects in prior work (e.g., Sacerdote, 2011; Epple and Romano, 2011).

Peer-adjusted school effects, by contrast, capture school inputs and practices that do not adjust in response to random year-to-year changes in peer composition. These include school facilities and infrastructure, the overall quality of teachers and school leadership, established pedagogical models, and organizational culture. This distinction differs from conventional school value-added models, in which estimated school effects generally combine these relatively stable features with effects operating through peer composition.[5]

3. Identification

Identifying causal school and peer effects requires addressing endogeneity in both school assignment and peer composition. If unobserved student ability is correlated with the school a student attends or the peers she encounters, naïve comparisons may attribute outcome differences to schools or peers when they instead reflect selection. A second challenge is that school and peer treatments are assigned jointly: each school assignment comes with a peer assignment, and each

[4] Variance equals the second raw moment minus the square of the first. The functional term can capture the former but not the latter.

[5] To illustrate, let $\bar{P}_{jt}$ denote the average peer contribution among students attending school j in year t. Holding the adjustment for student characteristics fixed and abstracting from residual student sorting and sampling error, a model that omits peer composition attributes $\mu_{jt} + \bar{P}_{jt}$ to the school, up to a common normalization.

peer environment is observed through a school assignment. Random variation in school assignment can therefore identify the value of being assigned to a school, but that value combines the school's own contribution with the contribution of the peers assigned there. Conversely, variation in realized peer composition alone may fail to identify peer effects if some peer environments are more likely to occur at schools with higher or lower peer-adjusted school effects. Separately identifying peer effects and peer-adjusted school effects therefore requires variation in peer composition that is not fully collinear with school assignment, as well as variation in school assignment itself. Prior work shows that school-choice lotteries generate random variation in school assignments that can be used to identify causal school effects (Abdulkadiroğlu et al., 2017). Our identification builds on this insight by showing that lotteries also generate distinct random variation in peer composition.

3.1 Algorithmic School Assignments and Identifying Variation

We study the identification of school and peer effects in a centralized school-assignment market in which each student is assigned to at most one school and no school admits more students than its capacity. Centralized school-choice systems are now widespread, with national systems in countries such as Chile, Hungary, and Romania and city-level systems in Boston, New York, Bogotá, and Mexico City, among others. A school assignment is summarized, for each student i, by a vector of indicators $d_i = (d_{i1}, \ldots, d_{iS})$, where d_{ij} equals one if student i is assigned to school j and zero otherwise. Given the full vector of assignment indicators, the distribution of relative achievement among student i's co-assigned peers is determined. We denote this distribution by $f_i^{\text{asg}}(r)$.

In these markets, students submit rank-ordered preferences over schools, and schools have fixed capacities. Students also receive a priority status at each school. For example, a school may give priority to applicants whose home is closer to the school. Let $\succ_i$ denote student i's submitted rank-ordered list, and let $\boldsymbol{\rho}_i$ collect her priority statuses across schools. We summarize this information by student i's type, $\theta_i = (\succ_i, \boldsymbol{\rho}_i)$.

Priorities are typically coarse: at a given school, many applicants may share the same priority status. Random lottery numbers resolve these ties. That is, among students in the same priority group, the lottery determines their relative order at the school, yielding a strict priority ordering.

Given students' submitted preferences, priority statuses, school capacities, and the realized lottery numbers, an assignment algorithm produces a feasible assignment of students to schools. As noted by Abdulkadiroğlu et al. (2017), before lottery numbers are realized, the same procedure induces a set of assignment probabilities, p_{ij}, for each student and school pair. We add that it also induces, for each student, an expected[6] co-assigned peer density, $\widetilde{f}_i^{\text{asg}}(r) = \mathbb{E}(f_i^{\text{asg}}(r)|\theta)$, with $\theta \equiv (\theta_{i'})_{i'}$, summarizing the expected relative-achievement distribution of students assigned to the same school as student i.

The central implication is that assignment propensity scores absorb the systematic component of selection into school assignments (Abdulkadiroğlu et al., 2017), while expected co-assigned peer densities absorb the systematic component of selection into co-assigned peer composition. Conditional on these expected assignment objects, realized school assignments and co-assigned peer compositions vary with random lottery draws rather than with unobserved student ability. Moreover, lottery-induced deviations in peer composition are orthogonal to each school's unobserved peer-adjusted effectiveness, providing variation that we use to identify peer effects separately from peer-adjusted school effects.

More generally, school assignments and co-assigned peer compositions are "formula instruments" for school enrollment and co-enrolled peer composition. They are determined by a known formula that combines random and nonrandom inputs, and adjusting for the expected instrument—where the expectation is taken over the random inputs, holding fixed the nonrandom inputs—can eliminate omitted-variable bias arising from selection (Borusyak and Hull, 2023). In our setting, the corresponding expected instruments are the school-assignment probabilities, p_{ij}, and the expected co-assigned peer density, $\widetilde{f}_i^{\text{asg}}(r)$.

To clarify how the two sources of random variation separately identify peer effects and peer-adjusted school effects, we present the argument in two steps. The first step uses assignment-induced variation in enrollment to identify the composite value of attending each school, combining the school's own contribution with the contribution of the peers enrolled there. The second step uses assignment-induced variation in peer composition to decompose this composite value into a peer component and a peer-adjusted school component. Our empirical specification in Section 5.3

[6]The expectation is taken over the distribution of lotteries.

implements a single-step estimator; the two-step argument is intended only to clarify the sources of identifying variation.

To simplify notation, we suppress the cohort subscript t and present the argument for a single assignment cohort. This is only a notational simplification: the argument does not rely on observing multiple cohorts or on assuming that peer-adjusted school effects are constant over time. With multiple cohorts, the same argument applies cohort by cohort, allowing peer-adjusted school effects to vary over time.

In the first step, we identify the composite value of attending each school. Fix a value of baseline achievement x and consider students with $x_i = x$. Let $f_j^{\text{enr},x}(r;\mathbf{s})$ denote the distribution of peer achievement relative to x among students enrolled in school j under enrollment vector $\mathbf{s}$. Allow the peer-adjusted school effect and the peer-effect function to be indexed by x. Define the composite value of enrolling in school j for students with baseline achievement x as

$$\delta_j^x(\mathbf{s}) = \int \beta^x(r) f_j^{\text{enr},x}(r;\mathbf{s})\, dr + \mu_j^x \,. \tag{2}$$

Then, the outcome equation for students with baseline achievement x can be written as $y_i = \sum_{j=1}^{J} s_{ij}\delta_j^x(\mathbf{s}) + \varepsilon_i$.

Using the assignment indicator d_{ij} and assignment propensity score p_{ij} defined above, define the school-assignment instrument as $z_{ij}^S = d_{ij} - p_{ij}$, where the superscript S denotes the school-assignment shock. Following the formula-instrument logic in Borusyak and Hull (2023), this instrument isolates the lottery-induced component of assignment. For students with $x_i = x$, and provided there is non-degenerate assignment risk across the relevant schools, these instruments identify the composite values $\delta_j^x(\mathbf{s})$ through the moment conditions

$$\mathbb{E}\left[z_{ij}^S \left(y_i - \sum_{\ell=1}^{J} s_{i\ell}\delta_\ell^x(\mathbf{s}) \right) \,\middle|\, x_i = x \right] = 0, \qquad j = 1, \ldots, J.$$

In the second step, we separate peer effects from peer-adjusted school effects using the identified composite school values. For a given baseline-achievement level x, consider the school-level decomposition in Equation (2). The peer-composition term may be endogenous because schools

with different peer-adjusted effects may systematically enroll different peer environments. That is, $f_j^{\text{enr},x}(r;\mathbf{s})$ may be correlated with μ_j^x. We therefore use the peer-composition analogue of the enrolled school instrument. Let $f_j^{\text{asg},x}(r)$ denote the relative-achievement distribution, for a student with baseline achievement x, of peers assigned to school j, and let $\widetilde{f}_j^{\text{asg},x}(r) = E(f_i^{\text{asg},x}(r)|\theta)$ denote the corresponding expected assigned-peer density. Define the peer-composition instrument as

$$z_j^P(r;x) = f_j^{\text{asg},x}(r) - \widetilde{f}_j^{\text{asg},x}(r),$$

where the superscript P denotes the peer-composition shock. Following the formula-instrument logic in Borusyak and Hull (2023), this instrument isolates the lottery-induced component of peer composition. For students with $x_i = x$, and provided there is non-degenerate peer-composition risk across the relevant schools, these instruments identify the peer-effect function through the moment condition

$$\mathbb{E}\left[z_j^P(\cdot;x)\left(\delta_j^x(\mathbf{s}) - \int \beta^x(r) f_j^{\text{enr},x}(r;\mathbf{s})\,dr\right) \,\middle|\, x_i = x \right] = 0.$$

Once $\beta^x(r)$ is identified, the peer-adjusted school effect μ_j^x is identified by subtracting the implied peer contribution from the composite school value $\delta_j^x(\mathbf{s})$.

The preceding argument establishes identification in a model that is more general than the one we estimate. In the baseline empirical specification, we impose the restrictions $\beta^x(r) = \beta(r)$ and $\mu_j^x = \mu_j$ for all x. We also consider specifications in which $\beta(r)$ and μ_j vary across discrete student types, most prominently two groups defined by baseline achievement.

The identification argument highlights a tension between the conditions needed for consistent estimation of peer-adjusted school effects and peer effects: consistent estimation of each school's composite value requires sufficiently many students per school. At the same time, as schools become larger, realized peer composition converges to its expected value, reducing the lottery-induced variation that identifies peer effects (Angrist, 2014). Appendix F characterizes this trade-off formally and derives the asymptotic conditions under which both can be consistently estimated. Consider a sequence of increasingly large school environments, indexed by n, in which the number of students N_n grows and the number of schools J_n may grow with it. Under stylized assumptions,

the appendix shows that consistent estimation of each school's composite value requires the number of students per school, N_n/J_n, to grow, so that $J_n/N_n \to 0$ as $n \to \infty$. Consistent estimation of peer effects, by contrast, requires sufficiently many, sufficiently small schools, yielding the condition $J_n/\sqrt{N_n} \to \infty$ as $n \to \infty$. Thus, both peer-adjusted school effects and peer effects can be consistently estimated when J_n grows faster than $\sqrt{N_n}$ but slower than N_n.

3.2 Practical Considerations: Rank Conditions and Noncompliance

One practical consideration in identifying our model is that some schools in our data are undersubscribed, implying they have few or no applicants for whom assignment risk is strictly between 0 and 1. In this case, risk-adjusted school-assignment offers are constant at zero and therefore uninformative about a school's causal effects. This limitation has been well recognized in the literature (Angrist et al., 2024). Following this approach, we address the undersubscription challenge by imposing conditional independence assumptions that require enrollment decisions and co-enrolled peer composition to be as good as random conditional on assignment risk and student covariates. These assumptions allow us to recover causal estimates of schools and peers under an OLS strategy that controls for the expected treatments. We next state these conditional independence assumptions explicitly and show how they lead to the OLS estimating equation used in the empirical analysis.

The first assumption follows the risk-controlled conditional independence condition in Angrist et al. (2024) and rules out selection through noncompliance.

Assumption 1 (Risk-Controlled CIA (RC-CIA))**.** *Conditional on student covariates and assignment propensity scores, enrolled school is independent of unobserved determinants of achievement:*

$$\varepsilon_i \perp s_i \mid (X_i, p_i).$$

Under perfect compliance, $s_i = d_i$ for all i, this condition follows mechanically from the randomization generated by the assignment mechanism.

The second assumption augments RC-CIA to include co-enrolled peer composition. It rules out selection into peer groups through coordinated or otherwise systematic noncompliance.

Assumption 2 (Augmented Risk-Controlled CIA (aRC-CIA))**.** *Conditional on student covariates, assignment propensity scores, and expected co-assigned peer composition, enrolled school and co-enrolled peer composition are independent of unobserved determinants of achievement:*

$$\varepsilon_i \perp (s_i, f_i^{\text{enr}}) \mid (X_i, p_i, \widetilde{f}_i^{\text{asg}}).$$

Relative to RC-CIA, the additional content of this assumption is that a student's co-enrolled peer distribution is independent of her unobserved determinants of achievement after conditioning on observed characteristics, assignment propensity scores, expected co-assigned peer composition, and enrolled school.[7] The assumption would fail if students' enrollment decisions generated co-enrolled peer groups that are systematically related to unobserved ability in ways not captured by X_i, p_i, and $\widetilde{f}_i^{\text{asg}}$. For example, families can coordinate to decline their assigned schools and enroll their children together elsewhere, with this coordination reflecting shared parental involvement or student motivation that also affects achievement. Such behavior could generate selection into peer groups beyond what is explained by observed family characteristics, students' prior achievement, and their expected school and peer assignments. As with RC-CIA, under perfect compliance, this condition follows from the randomization generated by the assignment mechanism.

The aRC-CIA assumption underlies our main OLS strategy in Section 5.3, which conditions on the expected treatments generated by the assignment mechanism: school-assignment propensity scores and expected co-assigned peer composition. Provided that school enrollment and co-enrolled peer composition retain sufficient variation after conditioning on these objects and student characteristics, aRC-CIA identifies peer effects and peer-adjusted school effects without requiring a 2SLS implementation.[8] The assignment-based instruments excluded from the estimating equation then provide additional moment restrictions, which we use to conduct an overidentification test of the orthogonality conditions implied by aRC-CIA.

[7] More specifically, aRC-CIA can be decomposed into the following two conditions:

$$\varepsilon_i \perp s_i \mid (X_i, p_i, \widetilde{f}_i^{\text{asg}}), \text{ and } \varepsilon_i \perp f_i^{\text{enr}} \mid (X_i, p_i, s_i, \widetilde{f}_i^{\text{asg}}).$$

[8] Section 5.4 empirically assesses how much variation remains in school enrollment and co-enrolled peer composition after conditioning on the expected treatments.

4. FPCA: Reducing the Dimension of Peer Distributions

Our model allows student achievement to depend on the entire distribution of peer prior achievement, f_i.[9] This flexibility is central to the analysis, but it poses a practical challenge: each peer distribution is a high-dimensional object that cannot be included directly in a conventional outcome equation. We address this challenge using Functional Principal Component Analysis (FPCA) following Chang et al. (2023). FPCA captures the leading variation in peer distributions with a small number of common basis functions and corresponding student-specific loadings. This representation allows us to replace f_i with a small set of regressors and to estimate the peer-effect term in Equation (1).

To implement FPCA, we evaluate each peer distribution f_i on a common grid of relative-achievement values $(r_1, \ldots, r_T)$ and represent it by the vector $\big(f_i(r_1), \ldots, f_i(r_T)\big)$. FPCA first centers the discretized peer distributions point by point. That is, at each value of relative achievement on the grid, it subtracts the average value of the peer distribution across students. It then applies the functional analogue of principal component analysis to the resulting deviations from this pointwise average. This procedure yields a set of basis functions that are common across students; we retain the first M, denoted by $v_1(r), \ldots, v_M(r)$. In the implementation, each basis function $v_m(r)$ is represented by its values on the same grid as the peer distributions: $\big(v_m(r_1), \ldots, v_m(r_T)\big)$. Formally, in the functional formulation, the basis functions are eigenfunctions of the covariance operator of the centered peer distributions. Their grid-based representations are the corresponding analogue of the eigenvectors of a covariance matrix in conventional principal component analysis.

The basis functions summarize the principal ways in which peer distributions differ across students. The first basis function captures the pattern of deviations from the average peer distribution that accounts for the greatest variation across students. Each subsequent basis function captures the largest remaining source of variation, subject to being orthogonal to the preceding functions. In the discretized representation, this means that the vector of values associated with one basis function is orthogonal to the corresponding vector for any other basis function. Thus, each component captures a separate pattern of variation in peer composition rather than repeating

[9] Throughout this section we use f_i generically to denote a peer distribution. In the empirical implementation, we apply FPCA separately to the enrolled and expected assigned peer distributions defined in the identification section.

variation already captured by an earlier component.

Each peer distribution can be approximated by the average peer distribution plus a linear combination of the first M basis functions:

$$f_i(r) \approx \bar{f}(r) + \sum_{m=1}^{M} c_{im} v_m(r), \tag{3}$$

where c_{im} is student i's loading on basis function m.[10] Intuitively, the loading c_{im} measures the extent to which student i's peer distribution exhibits the pattern of variation represented by $v_m(r)$. A loading with a larger absolute value indicates that student i's peer distribution departs more strongly from the average distribution along component m. Given the orientation of the basis function, the sign of the loading indicates the direction in which the distribution departs from the average along that pattern.[11]

FPCA chooses the loadings to provide the best approximation of each student's centered peer distribution using the retained basis functions. In the discrete implementation, c_{im} corresponds to the inner product between the centered vector of peer-distribution values for student i and the vector of grid values associated with basis function m. Substituting this representation into the peer-effect term in Equation (1) yields

$$\int \beta(r)\, f_i(r)\, dr \;\approx\; \int \beta(r)\, \bar{f}(r)\, dr \;+\; \sum_{m=1}^{M} c_{im} \int \beta(r)\, v_m(r)\, dr.$$

The first term is common across students and can therefore be absorbed by the constant in Equation (1). Define $\gamma_m \equiv \int \beta(r)\, v_m(r)\, dr$. The outcome equation can then be written as

$$y_{it} \;=\; \alpha \;+\; \mu_{jt} \;+\; \sum_{m=1}^{M} \gamma_m\, c_{im} \;+\; \epsilon_{it}.$$

Here, the scalar loadings $c_{i1}, \ldots, c_{iM}$ enter the outcome equation as regressors, while the coefficients $\gamma_1, \ldots, \gamma_M$ capture the contribution of each dimension of variation in peer composition

[10] Note that each peer distribution $f_i(r)$ integrates to one over its support, as does the average distribution $\bar{f}(r)$. Thus the centered distributions $f_i(r) - \bar{f}(r)$ integrate to zero. Since the FPCA basis functions are constructed from variation in these centered distributions, each basis function also integrates to zero.

[11] FPCA determines each component only up to sign. That is, $v_m(r)$ and $-v_m(r)$ represent the same component with the opposite orientation. Reversing the orientation also reverses the associated loadings. Thus, the sign of a loading is meaningful only relative to the chosen orientation of its basis function.

to student achievement. FPCA thus replaces the peer distribution with a small number of scalar regressors that can be incorporated into a standard linear regression. We discuss our choice of M, together with other implementation decisions, in Section 5.2.

These regression coefficients can then be used to recover the peer-effect function evaluated on the discrete grid used for estimation. Recall that γ_m measures the contribution to achievement of the peer-composition pattern represented by $v_m(r)$. We recover the estimated peer-effect function by combining the retained basis functions, weighting each one by its corresponding estimated regression coefficient. This translates the regression estimates back from the component representation to estimated peer effects at each relative-achievement value on the grid:

$$\hat{\beta}(r_t) = \sum_{m=1}^{M} \hat{\gamma}_m v_m(r_t), \qquad \text{for each } \; t = 1, \ldots, T.$$

The resulting vector $(\hat{\beta}(r_1), \ldots, \hat{\beta}(r_T))$ is the discrete approximation to the peer-effect function implied by the retained basis functions. As shown in Chang et al. (2023), under standard regularity conditions, $\hat{\beta}$ consistently estimates the projection of β onto the space spanned by the retained principal components.

This approach has two useful properties for our setting. First, FPCA provides a disciplined way to reduce the dimensionality of peer composition. Among all M-dimensional linear representations of the peer distributions, the first M principal components minimize average squared approximation error (Chang et al., 2023). Thus, for a given number of regressors, FPCA retains as much of the cross-student variation in peer composition as possible. This is important because our objective is to capture as much variation in peer composition as possible while keeping the empirical specification tractable.

Second, FPCA produces mutually uncorrelated loadings, avoiding variance inflation from multicollinearity among the retained FPCA regressors and thereby improving the precision of the estimated coefficients. Moreover, the retained loadings are orthogonal to the omitted higher-order components. Consequently, truncating the FPCA representation does not mechanically bias the coefficients on the retained components. Appendix D provides additional details on these properties.

These advantages should be interpreted with one qualification. FPCA constructs components to capture the dominant dimensions of cross-student variation in peer distributions, not the dimensions that are most predictive of outcomes. Thus, the leading components are not necessarily those most relevant for achievement. We nevertheless leverage this approximation because of its favorable finite-sample properties: even if we could identify features of peer distributions that are particularly relevant for outcomes, limited sample variation along those directions could yield imprecise and potentially uninformative estimates. As a robustness exercise, we also consider alternative representations that more directly target estimation of the peer effect function, such as B-splines, and augment the FPCA representation with peer characteristics that theory suggests may be particularly relevant for achievement. These exercises allow us to assess the sensitivity of our estimates to the particular representation of peer composition.

5. Empirical Setting and Implementation

We next describe the institutional setting and data, and the details of our implementation of the FPCA procedure and subsequent regression analysis.

5.1 Institutional Setting and Data

We use administrative data from the New York City Department of Education (DOE) to construct peer distributions, measure student outcomes, and observe the school-assignment and school-enrollment variables used in our identification strategy. We use data for three middle-school application cohorts: fifth-grade students applying for sixth-grade admission in 2014–15, 2015–16, and 2016–17.[12] Together, these cohorts include 216,525 students. For each student, we observe submitted rank-ordered lists, final school assignments, enrollment decisions, and admissions priorities at each school program. We also observe New York State Math and English Language Arts (ELA) test scores, as well as demographic characteristics including race and ethnicity, English Language Learner status, and eligibility for subsidized lunch. Among all students in these cohorts, 195,539 have non-missing baseline characteristics; we treat this group as the population sample in Table 1.

Panel A of Table 1 reports summary statistics for student characteristics. After the cleaning

[12] See Appendix E.1 for a description of the admissions process.

Table 1: Student and School-by-Cohort Summary Statistics

	(1) Full sample	(2) Analysis sample
Panel A. Student-level sample and characteristics		
Students	195,539	143,147
2014 cohort	63,838	46,601
2015 cohort	65,430	48,094
2016 cohort	66,271	48,452
Mean baseline score	0.022	0.053
	(0.987)	(0.986)
Share male	0.507	0.504
Share Asian	0.174	0.190
Share Black	0.233	0.219
Share Hispanic	0.418	0.417
Share free/reduced-price lunch	0.714	0.714
Share special education	0.214	0.199
Share English learner	0.124	0.110
Panel B. School-by-cohort characteristics		
School-by-cohort observations	1,476	1,473
2014 cohort	503	500
2015 cohort	490	490
2016 cohort	483	483
Unique schools	507	507
Mean sixth-grade cohort size	117	97
	(109)	(93)
Share using random tie-breaker	0.753	0.754

Notes: This table reports summary statistics for the full sample and the analysis sample. The full sample consists of applicants for a sixth-grade seat with non-missing baseline characteristics. The analysis sample further restricts to students with observed school assignment, observed school enrollment, and non-missing sixth-grade test scores. Application cohorts refer to fifth-grade students applying for sixth-grade admission. Baseline test scores are the average of standardized fifth-grade Math and ELA scores. Shares are reported as fractions. Standard deviations for continuous variables are reported in parentheses. School-by-cohort observations are defined at the school-by-sixth-grade-entry-cohort level. The random tie-breaker row reports the share of school-by-cohort observations in which at least one program in the school used a random tie-breaker.

steps described in Appendix E.2, the analysis sample contains 143,147 students. Most excluded observations are missing either assigned or enrolled schools: 49,982 of the 62,831 dropped observations fall into this category. The analysis and full samples have similar demographic compositions. In both columns (1) and (2), half of the students in the analysis sample are male; 22 percent are Black, 42 percent are Hispanic, and 19 percent are Asian. The mean baseline test score, defined as the average of standardized Math and ELA scores, is modestly higher in the analysis sample (0.053 versus 0.022), indicating that students in the estimation sample are slightly higher-achieving than the full population of NYC students with baseline characteristics.

Our analysis defines peer groups at the school-by-cohort level, where a cohort is the set of students entering sixth grade in the same year. This choice aligns our measurement of peer composition with our objective of comparing school and peer contributions to achievement and with our optimal assignment exercise, which considers alternative allocations of students across schools. Although schools may offer multiple programs, and programs are the relevant unit for assignment, students assigned through different programs within the same school share common spaces, extracurricular activities, and classes. The full sample contains 1,476 school-by-cohort observations, and the analysis sample contains 1,473.

The school, rather than the program, is also the unit at which we define school effects. There are 509 distinct schools in the full sample and 507 in the analysis sample. The average entering sixth-grade cohort in the analysis sample contains 97 students, with a standard deviation of 93, reflecting substantial variation in cohort size across schools.

Random tie-breakers are used in 75 percent of school-by-cohort observations in the analysis sample. The remaining observations correspond to school cohorts whose programs use screened or selective admissions methods, which construct applicant rankings using non-lottery criteria, such as grades, attendance, auditions, interviews, or school-specific composite rules. Importantly, the absence of a random tie-breaker at a given program does not imply an absence of quasi-random variation in school assignment or peer composition. Because students submit ranked application lists, random rejection from a program that uses a random tie-breaker can induce quasi-random assignment to another program that does not. Table 2 describes this variation in more detail.

5.2 Constructing Peer Distributions and Implementing FPCA

We measure peer quality relative to the focal student's own prior achievement. Two students are defined as peers if they enroll in the same school and enter sixth grade in the same year. For each peer k of student i, relative achievement is $r = y_{k,t-1} - y_{i,t-1}$, the difference between peer k's lagged achievement and student i's lagged achievement. The peer distribution $f_i^{\text{enr}}(r)$ is therefore the distribution of these relative-achievement values across student i's co-enrolled peers. In our implementation, we estimate $f_i^{\text{enr}}(r)$ by applying kernel density smoothing to student i's set of relative-achievement values and evaluating the resulting density on a common, equally spaced 100-point grid.[13]

Since the tails of each peer distribution have little variation, we restrict the support to the central 95 percent of the pooled enrolled-peer density and redistribute each student's omitted tail mass back over the retained support using boundary-concentrated normal weights.[14] The normal weights spread the omitted tail mass over multiple points of the retained grid rather than creating mass points at the boundaries, concentrating lower-tail mass near the lower boundary and upper-tail mass near the upper boundary.

We then apply the FPCA procedure described in Section 4 to these peer distributions. This yields a grid-based decomposition of variation in peer distributions, with each of the 100 basis functions represented by a vector containing one value at each point of the retained grid. In our application, we retain the first three basis functions; that is, we set $M = 3$.

Appendix Figure A.1 summarizes the information used to guide this choice. Figure A.1a reports the share of variation in peer distributions explained by the leading principal components, computed relative to the total variation captured by the full FPCA decomposition. The first three components explain 94.5% of this variation. Figures A.1b and A.1c show how out-of-sample prediction of student achievement changes with the number of components used to approximate peer distributions. These prediction exercises show that the outcome-relevant information in peer-achievement distributions is concentrated in the first few FPCs, and suggest that components

[13] We implement this step using MATLAB's `ksdensity` function evaluated on the common grid, with the default bandwidth selection and consider robustness to grid and bandwidth choices.

[14] Where peer densities vary little across students, the retained FPCA basis functions tend to be close to zero, mechanically pushing the reconstructed peer-effect function toward zero in those regions. This reflects limited variation in peer exposure rather than evidence that peer effects are absent.

beyond the first three add little useful out-of-sample predictive content. We also show that our main results are robust to including a fourth component (Figure A.10b).

The three retained basis functions summarize the main ways in which peer-achievement distributions differ across students (Appendix Figure A.2). The first component captures whether peers generally have higher or lower prior achievement than the focal student, with higher loadings indicating greater exposure to higher-achieving peers. This component shows that the main dimension along which peer distributions differ is whether students are exposed to peers who are, on average, above or below them in prior achievement. The second contrasts distributions concentrated near the focal student's own achievement with those having more mass in the tails, especially the upper tail. The third captures more localized differences in peer composition, particularly in the prevalence of peers whose achievement is slightly below or far above the focal student's own.

Having selected the first three basis functions, we next use them to construct the regressors that enter the achievement model. In particular, we use these components to summarize each student's co-enrolled peer distribution, $f_i^{\text{enr}}(r)$ as well as expected co-assigned peer distribution, $\widetilde{f}_i^{\text{asg}}(r)$. We now describe how these FPCA summaries enter the estimating equation.

5.3 Estimation

With the FPCA representation of peer distributions in hand, we now estimate the achievement model. Let c_{im}^{enr} denote student i's loading on basis function m for the co-enrolled peer distribution f_i^{enr}. Let $\widetilde{c}_{im}^{\text{asg}}$ denote the corresponding loading for the expected co-assigned peer distribution $\widetilde{f}_i^{\text{asg}}(r)$. Thus, c_{im}^{enr} summarizes the realized peer composition in the school where student i enrolls, while $\widetilde{c}_{im}^{\text{asg}}$ summarizes the expected peer composition under the assignment mechanism.

Starting from Equation (1) and building on the identification strategy described in Section 3, we estimate the following specification by OLS:

$$y_i = \sum_{j=1}^{J} \mu_j s_{ij} + \sum_{m=1}^{M} \gamma_m c_{im}^{\text{enr}} + \sum_{j=1}^{J} \kappa_j p_{ij} + \sum_{m=1}^{M} \rho_m \widetilde{c}_{im}^{\text{asg}} + X_i'\alpha + \epsilon_i. \tag{4}$$

Recall that s_{ij} is an indicator for whether student i enrolls in school j, and p_{ij} is the assignment propensity score: the probability that student i is assigned to school j under the centralized

assignment mechanism. The vector X_i includes student baseline characteristics: fifth-grade Math and ELA scores, their squares and cubes, and indicators for English Language Learner status, gender, race and ethnicity, free- or reduced-price lunch eligibility, and special education status. The coefficients μ_j capture peer-adjusted school effects, which we assume are time-invariant in our main specification, while the coefficients γ_m capture the contribution of peers to achievement. The terms p_{ij} and $\widetilde{c}_{im}^{\text{asg}}$ control for expected school assignment and expected co-assigned peer composition under the lottery mechanism.

We estimate Equation (4) for the full sample of students and also examine heterogeneity along several student characteristics. Our main heterogeneity analysis separates students by baseline achievement. We define low-achieving students as those with baseline achievement at or below the cohort-specific median and high-achieving students as those with baseline achievement above the median. Peer effects and peer-adjusted school effects for low- and high-achieving students are estimated in a single regression by interacting both the school-enrollment indicators, s_{ij}, and the co-enrolled peer-composition loadings, c_{im}^{enr}, with indicators for the two achievement groups, while restricting the remaining coefficients to be common across groups. In additional analyses, we consider other dimensions of heterogeneity, including achievement quartiles, gender, income, and race and ethnicity. The main results below report estimates for the full sample and for the two baseline-achievement groups.

After estimating Equation (4), we form empirical-Bayes posterior predictions of the school-effect parameters using linear shrinkage, following Walters (2024). In the heterogeneous specification, we shrink the school-specific intercepts and high-achieving-student interactions jointly, preserving their estimated covariance structure.

To conduct inference on the peer-effect function, we estimate Equation (4) with robust standard errors. We then use the resulting variance-covariance matrix of estimates to translate the estimated sampling uncertainty in the FPCA coefficients into uncertainty over $\widehat{\beta}(r)$. Let $\widehat{\gamma} = (\widehat{\gamma}_1, \ldots, \widehat{\gamma}_M)'$ denote the estimated coefficients on the co-enrolled peer-composition loadings, and let $\widehat{V}_\gamma$ denote the corresponding block of the market-clustered variance-covariance matrix. We construct pointwise confidence bands using a simulation based on the asymptotic distribution of $\widehat{\gamma}$. Specifically, we draw 10,000 coefficient vectors from the multivariate normal distribution $N(\widehat{\gamma}, \widehat{V}_\gamma)$ and, for each

draw, compute the implied peer-effect function. At each value of r on the estimation grid, we take the 2.5th and 97.5th percentiles of the simulated values. These percentiles form the pointwise 95 percent confidence band around $\widehat{\beta}(r)$. These confidence bands do not account for uncertainty from estimating the FPCA basis. To assess its stability, we re-estimate the co-enrolled-density FPCs in 100 school-cluster bootstrap samples and recalculate loadings for the original full sample using each bootstrap basis. The resulting basis functions vary little across bootstrap samples (Appendix Figure A.3). For each of the three retained components, the correlation between the recalculated loadings and their full-sample counterparts exceeds 0.999 in every bootstrap replication.

5.4 Lottery-induced Variation In Practice

Before turning to the estimates, we assess the empirical variation that underlies our research design. First, we document the extent of lottery-induced variation in school assignment and peer composition in our data. Second, we assess the identifying assumption by testing whether this lottery-induced variation in school assignment and peer composition is uncorrelated with achievement differences left unexplained by the model.

Table 2 shows that the assignment mechanism generates relevant random variation in both school assignments and peer composition. Panel A reports lottery-induced variation in school assignment. A student faces assignment risk when her probability of assignment to a given school lies strictly between zero and one. Among the 143,147 students in the analysis sample, 38% have a nondegenerate assignment probability at their first-choice school, and 47.4% have a nondegenerate probability for at least one listed school. At the school-year level, an average of 40.8% of first-choice applicants and 34.7% of all applicants face assignment risk.

Panel B summarizes lottery-induced variation in peer composition using the total variation distance (TVD) between realized and expected co-assigned peer densities.[15] The average applicant-level TVD is 0.026, meaning that the realized peer distribution differs from its expectation by the equivalent of reallocating about 2.6 percent of peer-density mass. This variation is larger for students who face assignment risk at some listed school than for those who do not: 0.039 compared with 0.015. Appendix Figure A.4 shows that this variation extends throughout the peer-

[15]For two peer-density functions f and g, total variation distance is defined as $\text{TVD}(f, g) = \frac{1}{2}\int |f(r) - g(r)|\, dr$. In our implementation, we compare the co-assigned peer density with its corresponding expectation.

Table 2: Variation in School Assignment and Peer Composition

	(1)	(2)	(3)	(4)	(5)	(6)
	N	Mean	SD	p25	p50	p75
Panel A. Variation in School Assignment: Assignment Probability $\in (0, 1)$						
Applicants	143,147					
At risk at the first-choice school		0.380				
At risk at any school on the list		0.474				
School-years	1,457					
Share of first-choice applicants at risk		0.408	0.327	0.111	0.348	0.674
Share of listed applicants at risk		0.347	0.238	0.152	0.337	0.495
Panel B. Variation in Peer Composition: TVD between Assigned and Expected Peer Densities						
Applicants						
All	143,147	0.026	0.040	0.006	0.014	0.029
– at risk at any listed school	67,832	0.039	0.050	0.009	0.021	0.049
– at risk at none of listed school	75,315	0.015	0.022	0.005	0.011	0.019
School-years						
All	1,457	0.023	0.027	0.009	0.018	0.031
– oversubscribed lottery	697	0.024	0.023	0.010	0.020	0.033
– undersubscribed or non-lottery	705	0.022	0.031	0.008	0.017	0.029

Notes: Panel A reports lottery-induced assignment risk from the DA propensity-score file. The applicant rows report the share of students whose assignment propensity score is strictly between zero and one for their first-choice school or for at least one listed school. The school-year rows summarize the corresponding shares across analysis-sample school-years. Panel B reports TVD between realized and expected peer densities. TVD is defined as one half of the integrated absolute difference between the two densities. Applicant-level TVD compares each student's realized assigned-peer density to her expected assigned-peer density from simulated assignment draws. School-year TVD compares the realized assigned-student density in a school-year to the corresponding expected density across simulated assignment draws. The analysis-sample restriction is applied when summarizing the statistics; expected densities are computed from the full available density sample before this final reporting restriction.

achievement distribution. The average distance between realized and expected densities is about 4% of expected density near the center of the distribution and rises to roughly 11–14% toward the tails.

Panel B also reports peer-composition variation at the school-by-cohort level. As expected, oversubscribed schools that use lotteries exhibit variation in assigned peer composition, with an average TVD of 0.024. Yet variation is nearly as large at schools that are undersubscribed or do not use lotteries, where the average TVD is 0.022. Thus, lottery-induced variation in peer composition extends beyond schools that directly use lotteries to allocate seats.

Because our OLS specification uses enrolled schools and co-enrolled peer composition, while lotteries randomize assignment, its causal interpretation relies on aRC-CIA. This assumption requires that, conditional on student characteristics, school-assignment probabilities, and expected co-assigned peer composition, enrolled schools and peers be unrelated to unobserved determinants of achievement. We assess this restriction by extending the validation procedure in Angrist et al. (2017, 2024) to accommodate both school and peer effects. We construct school-assignment shocks as realized assignment indicators minus assignment probabilities.[16] Peer-assignment shocks are deviations of realized co-assigned peer composition from its expectation, summarized using the same FPCA basis as in the main analysis. Under aRC-CIA and the restrictions of our empirical specification, these shocks should be orthogonal to the regression error. We estimate Equation (4) on the full analysis sample, obtain the residuals, and conduct an omnibus test of whether the assignment shocks jointly predict them, conditional on the controls, among students facing meaningful school-assignment and peer-composition risk.

The omnibus statistic can be decomposed into forecast-bias and overidentification components, as in Angrist et al. (2017, 2024). The forecast-bias component tests whether the model's predicted school-and-peer contributions are correctly scaled along the lottery-induced variation: if so, the forecast coefficient should equal one. The overidentification component asks whether the different instruments imply a common forecast coefficient, without requiring that coefficient to equal one. It therefore captures discrepancies across lottery-induced sources of variation that cannot be resolved

[16] Following Angrist et al. (2017, 2024), we aggregate school-assignment shocks within ventiles of estimated school value-added to limit the number of potentially weak instruments.

by a common rescaling of the model's predictions. Because our instruments include both school- and peer-assignment shocks, we further allocate the overidentification statistic between the two using a Shapley decomposition. Appendix G provides the formal derivation and implementation details.

Table 3 shows that the specification tests improve markedly as the controls required by our identification strategy are introduced. In Panel A, the forecast coefficient is not different from 1 once we add assignment propensity scores. In Panel B, the uncontrolled and observable-only specifications reject the omnibus null, whereas the specification with assignment propensity scores does not ($p = 0.872$). Adding expected peer composition—the preferred specification—also yields no rejection: the omnibus-test p-value is 0.328, and neither the forecast-bias nor the overidentification component is statistically significant.

As in Angrist et al. (2017, 2024), our procedure jointly tests the identifying assumption and the structural model specification. Thus, the null may be rejected either because aRC-CIA fails or because the structural model omits relevant heterogeneity in school or peer effects. Our setting introduces an additional potential source of rejection: the finite-dimensional approximation to peer distributions may be inadequate. A rejection therefore would not, by itself, identify which of these restrictions fails. To assess whether the results are driven by a particular functional-form or heterogeneity restriction, Appendix Table G.1 replaces the functional peer measure with the peer mean, retains four rather than three FPCA components, and allows school and peer effects to differ between students with high and low baseline achievement. All results support the joint restrictions underlying the empirical specification.

Table 3: Validation Tests for Forecast Bias and Overidentification

	(1) Uncontrolled	(2) +Observable	(3) +PS	(4) +Expected Peer
Panel A: Forecast Coefficient				
Coefficient	0.984	0.931	1.023	1.038
	(0.003)	(0.037)	(0.045)	(0.044)
First-stage F-statistic	5984.87	390.05	426.48	431.80
Panel B: Omnibus tests				
Forecast bias	26.17	3.46	0.28	0.75
	(0.000)	(0.063)	(0.599)	(0.387)
+ Overidentification (χ^2 d.f.)	235.91 (22)	35.98 (22)	15.33 (22)	24.70 (22)
	(0.000)	(0.031)	(0.848)	(0.312)
= Omnibus test (χ^2 d.f.)	262.08 (23)	39.44 (23)	15.60 (23)	25.45 (23)
	(0.000)	(0.018)	(0.872)	(0.328)
Panel C: Overidentification decomposition				
School effect	148.46	30.97	13.55	18.55
Peer effect	87.44	5.01	1.78	6.15
N (testing)	43,007	43,007	43,007	43,007
N (full model)	143,147	143,147	143,147	143,147

Notes: The table reports results from identification assumption tests, explained in detail in Appendix G. Each column estimates the model named in the column header (variations of Equation (4)), constructs school-VA ventiles from that same model, and tests whether residuals are orthogonal to lottery-induced school-assignment and peer-assignment shocks. Column (1) includes school fixed effects and three co-enrolled-peer FPCA loadings. Column (2) adds cohort fixed effects and student baseline characteristics; column (3) also includes school-assignment propensity scores; and column (4) further adds expected assigned-peer FPCA coordinates. The testing sample is restricted to students with nondegenerate school-assignment risk to any school on their list and whose TVD between assigned and expected peer densities is above the median in Table 2. Admission probabilities to each school are summed within each ventile to which the school is assigned, based on shrunken school VA estimates. The omnibus statistic is the aggregate LM-style test; the forecast-bias statistic tests whether the coefficient on the fitted outcome variable equals one when regressing observed outcome on the fitted outcome; and the overidentification statistic tests whether the assignment shocks explain any residual variation in the outcome variable after allowing the fitted values to be rescaled by the forecast coefficient. The decomposition allocates the overidentification component between school and peer shocks by conducting a Shapley decomposition.

6. Results

This section presents estimates of peer effects and peer-adjusted school effects from Equation (4). We begin with estimates for the full sample. We then examine heterogeneity by students' baseline achievement, splitting students into low- and high-achieving groups based on their cohort-specific median baseline test score.

For both the full-sample and heterogeneous specifications, we also estimate a linear-in-means version of the model. In this specification, we replace the flexible FPCA peer-effect term, $\sum_{m=1}^{M} \gamma_m c_{im}^{\text{enr}}$, with $b\bar{r}_i^{\text{enr}}$, where $\bar{r}_i^{\text{enr}}$ is the mean relative achievement of student i's co-enrolled peers. We analogously replace the FPCA control for expected peer composition, $\sum_{m=1}^{M} \rho_m c_{im}^{\text{asg}}$, with the expected mean relative peer achievement under the assignment mechanism. We report results from both specifications to compare the flexible peer-effect function with the more restrictive linear-in-means model.

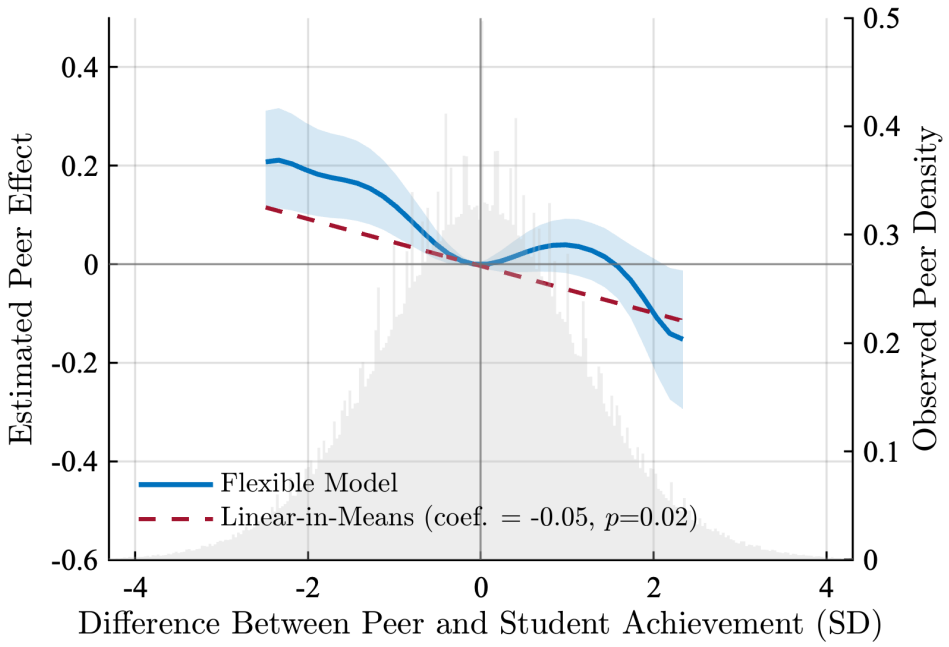
6.1 Peer Effects

Figure 1 reports the full-sample estimate of the peer-effect function. The horizontal axis measures peer achievement relative to the focal student's own baseline achievement, so moving to the right corresponds to shifting peer mass toward higher-achieving peers. The solid blue curve shows the flexible FPCA estimate, while the dashed red line reports the corresponding linear-in-means specification. Recall that the peer-effect function $\beta(r)$ is identified up to an additive constant. As a result, the level of the function at a particular value of r is not directly interpretable; instead, the relevant objects are differences such as $\beta(r_1) - \beta(r_2)$. In this and the following figures, we normalize both the flexible peer-effect function and the linear-in-means benchmark to equal zero at relative achievement $r = 0$. Thus, both lines pass through zero when peer achievement equals the focal student's own baseline achievement.

The estimated flexible peer-effect function is broadly downward sloping over the support of the peer distribution. Because differences in the value of the function determine the effect of reallocating peer mass across the distribution, this pattern implies that shifting peer mass toward higher-achieving peers results in lower achievement for the focal student. The confidence band lies above zero for negative values of relative peer achievement. Given the normalization discussed

above, this implies that replacing lower-achieving peers with same-achieving peers has a negative effect on achievement that is statistically distinct from zero. For positive values of relative peer achievement, the pattern is more mixed. The curve is slightly positive for peers just above the focal student, but declines and becomes negative for sufficiently higher-achieving peers, with confidence bands that include zero over much of this range. The flexible estimate broadly tracks the linear-in-means benchmark but reveals modest departures from a constant slope.

Figure 1: Estimated Peer Effects: Full Sample



Notes: This figure plots full sample estimates of the peer-effect function using Equation (4) for the relative-achievement specification, pooling the 2014–17 cohorts. Relative achievement is defined as a peer's baseline Math/ELA score minus the focal student's baseline Math/ELA score. The solid line reports the flexible FPCA estimate, reconstructed from the first three functional principal components of co-enrolled peer distributions. The dashed line reports the corresponding linear-in-means specification, which uses mean relative peer achievement. Both models control for expected peer composition, assignment propensity scores, year fixed effects, cubic polynomials in baseline Math and ELA scores, student demographics, and school fixed effects. The blue shaded region is a pointwise 95 percent confidence band computed from the market-clustered coefficient covariance matrix. The gray histogram shows the observed leave-one-out distribution of relative achievement among co-enrolled peers.

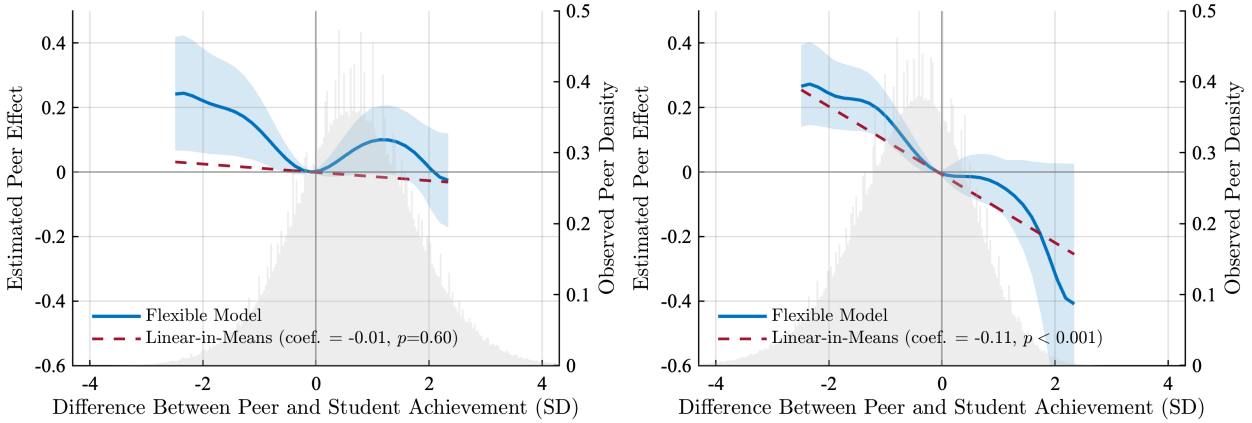
To summarize magnitudes, we compute the predicted effect of shifting each student's observed relative peer-achievement distribution by 0.5 standard deviations. This shift is comparable to the observed dispersion in peer exposure across schools: the standard deviation of school-specific mean baseline achievement is 0.5. Using the flexible estimate, shifting peer mass 0.5 standard deviations to the right lowers predicted achievement by about 0.029 SD. A shift of the same

magnitude to the left raises predicted achievement by a similar amount (-0.031 SD, Appendix Table B.1). The linear-in-means model implies symmetric effects by construction: increasing relative peer achievement by 0.5 standard deviations lowers predicted achievement by 0.024 SD, while decreasing it by the same amount raises predicted achievement by 0.024 SD.

The full-sample estimates mask heterogeneity by baseline achievement. Figure 2 reports estimates separately for low- and high-achieving students. For low-achieving students, shown in Panel (a), the flexible peer-effect function exhibits a nonlinear pattern: The peer effect function and the confidence band lie above zero for negative values of relative peer achievement. This implies that replacing lower-achieving peers with same-achieving peers has a negative effect on achievement that is statistically distinct from zero. To the right of zero, the function is nonmonotonic. It first increases and then declines at higher values of relative peer achievement, with normalized estimates that remain close to zero. The linear-in-means benchmark, by contrast, summarizes these offsetting patterns as an approximately zero peer effect. Using the same one-half-standard-deviation shift in relative peer achievement, the flexible estimates imply changes in predicted achievement of -0.01 SD for a rightward shift and 0.015 SD for a leftward shift (Appendix Table B.1).

For high-achieving students, shown in Panel (b) of Figure 2, the estimated peer-effect function is steeper and more systematically downward sloping. Moving to the right in the relative-achievement distribution is associated with lower achievement, especially for peers who are substantially higher-achieving than the focal student; confidence bands are wider in this region, so these upper-tail estimates should be interpreted more cautiously. The pointwise confidence band lies above zero throughout the negative relative-achievement range. This implies that lower-achieving peers raise predicted achievement relative to same-achieving peers. The linear-in-means model provides a close summary of the broad downward pattern for this group. Using the same one-half-standard-deviation shift in relative peer achievement, the implied changes are larger for high-achieving students than for low-achieving students in both the flexible and linear-in-means models. Under the flexible model, shifting the peer distribution to the right lowers predicted achievement for high-achieving students by 0.055 SD, while shifting it to the left raises predicted achievement by 0.052 SD (Appendix Table B.1). Thus, although negative peer effects are present for

Figure 2: Estimated Peer Effects: By Baseline Achievement



(a) Low-Achieving Students

(b) High-Achieving Students

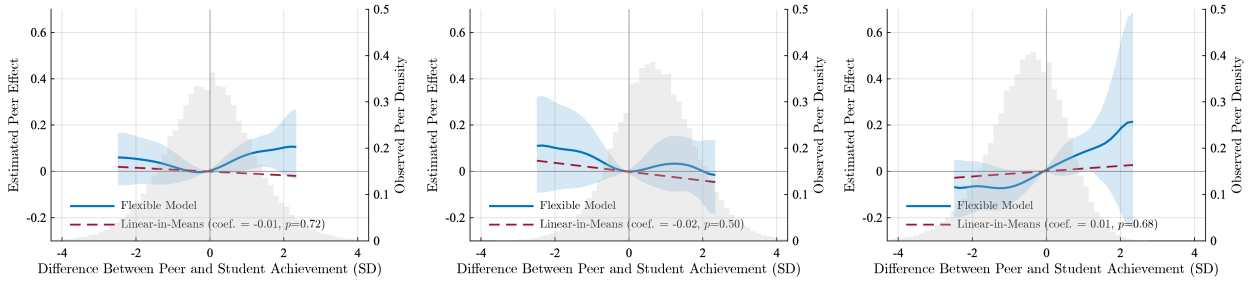
Notes: This figure plots estimates of the peer-effect function by baseline achievement group for the relative-achievement specification using Equation (4), pooling the 2014–17 cohorts. Baseline achievement is defined as the average of a student's standardized baseline Math and ELA scores. Low-achieving students are those with baseline achievement at or below the year-specific median; high-achieving students are those with baseline achievement above the year-specific median. Relative achievement is defined as a peer's baseline achievement minus the focal student's baseline achievement. The solid line reports the flexible FPCA estimate, reconstructed from the first three functional principal components of co-enrolled peer distributions. The dashed line reports the corresponding linear-in-means specification, which uses mean relative peer achievement. Both models allow school effects and peer effects to differ by baseline achievement group, while keeping the remaining coefficients common across groups. The blue shaded region is a pointwise 95 percent confidence band computed from the market-clustered coefficient covariance matrix. The gray histogram shows the observed leave-one-out distribution of relative achievement among co-enrolled peers.

both groups, the estimates for high-achieving students suggest a more pronounced and systematic pattern.

We find that the shape and the sign of peer effects may vary with the outcome considered. Figure 3 reports analogous estimates for a longer-term outcome, college enrollment. In the full sample (Panel (a)), the flexible peer-effect function rises for peers who are higher-achieving than the focal student, showing suggestive evidence that exposure to sufficiently higher-achieving peers raises predicted college enrollment. For low-achieving students (Panel (b)), the pointwise confidence band of the estimate includes zero throughout the support, showing muted peer effects on college enrollment for these students. For high-achieving students (Panel (c)), the estimated function is broadly upward sloping, although the confidence bands are wide, particularly for peers who are substantially higher achieving than the focal student. Again, to summarize magnitudes, for high-achieving students, the flexible model predicts that shifting the relative peer-achievement

distribution by 0.5 standard deviations to the right increases college enrollment by 2.89 percentage points, while an equally sized leftward shift decreases enrollment by 2.17 percentage points. Taken together, these estimates hint, if anything, at small but positive effects of exposure to higher-achieving peers on college enrollment, in contrast to the overall negative effects on sixth-grade achievement.

Figure 3: Peer Effects on College Enrollment



(a) Full Sample (b) Low-Achieving Students (c) High-Achieving Students

Notes: This figure plots estimates of the peer-effect function on college enrollment for the relative-achievement specification using Equation (4), pooling the 2014–17 cohorts. The outcome is measured in 0–1 units. The rest of the specification follows that of Figures 1 and 2.

Appendix Figure A.5 extends the analysis to seventh- and eighth-grade achievement, SAT scores, high school graduation, and college destinations. Consistent with the sixth-grade results, the estimated peer-effect functions for test-score outcomes, including the SAT, suggest a negative peer effect. The estimates for high school graduation and college destinations are less precise but suggest patterns similar to those for college enrollment. For high school graduation, the flexible estimate is relatively flat over much of the support and rises for peers who are sufficiently higher-achieving than the focal student. The estimated functions for enrollment in out-of-state colleges and private colleges in New York State are upward sloping overall. More broadly, these patterns are consistent with Boucher et al. (2024), who also find that peer effects differ across adolescent outcomes.

6.2 Peer-Adjusted School Effects

Table 4 reports noise-corrected estimates of the distribution of school-effect parameters. Panel A, column (1), reports estimates from Equation (4) under homogeneous school and peer effects. The

cross-school standard deviation of peer-adjusted school effects is 0.16σ. As a benchmark, this magnitude is close to the lower end of estimates of middle-school value added in other settings.[17] The heterogeneous specification in columns (2) and (3) shows evidence that school effectiveness varies with student baseline achievement after accounting for peer influence. The estimated cross-school standard deviation of the coefficient on the high-achieving-student indicator is 0.11σ. This implies that, holding a school's effect on low-achieving students fixed, a one standard deviation increase in its relative effectiveness for high-achieving students raises achievement for these students by 0.11σ relative to low-achieving students. Combining the estimated standard deviations of the school effect for low-achieving students and the high-achieving-student interaction, together with their estimated correlation, the implied cross-school standard deviation of peer-adjusted school effects is 0.196σ for high-achieving students, slightly larger than the corresponding standard deviation for low-achieving students, 0.169σ.[18] At the same time, school-effect estimates for low- and high-achieving students are highly aligned (slope = 1.02, SE = 0.02; $R^2 = 0.82$), suggesting broadly similar differences in school effectiveness across the two achievement groups (Figure A.6).

A natural follow-up question is how accounting for peer composition changes estimated school effects. This comparison connects our analysis to the large literature on school value added, in which estimated school effects capture both peer- and peer-adjusted school effects (see Angrist et al., 2023, for a review). To begin with, the measured dispersion in school effects does not substantially change when peer composition is accounted for. Panel B of Table 4 reports school effect estimates from an analogous model that does not account for peer composition.[19] The standard deviation of school effects for low-achieving students is nearly unchanged, at 0.166σ, while the implied standard deviation for high-achieving students is slightly lower than in the peer-adjusted model, at 0.180σ. The correlation between the school effect for low-achieving students and the high-achieving-student interaction is negative and statistically different from zero in the school-effect-only model, but small and statistically indistinguishable from zero in the peer-adjusted model. This pattern

[17] Angrist et al. (2017) find that the standard deviation of school value added across Boston middle schools is 0.15–0.20σ.

[18] The school effect for high-achieving students is the sum of the school effect for low-achieving students and the high-achieving-student interaction. Using the noise-corrected standard deviations and correlation in Table 4, Panel A, its standard deviation is $\sqrt{0.169^2 + 0.112^2 + 2(-0.066)(0.169)(0.112)}\,\sigma \approx 0.196\sigma$.

[19] This model is analogous to the benchmark model in Equation (4), but omits the co-enrolled peer-composition loadings and the controls for expected peer composition.

Table 4: Moments of Peer-Adjusted School Effects and School-Effects

	(1)	(2)	(3)
	Homogenous	Heterogeneous	
		Intercept	High-Achieving Students
Panel A: Peer-Adjusted School Effects			
Mean	0	0	0.068
	-	-	(0.065)
SD	0.162	0.169	0.112
	(0.006)	(0.006)	(0.008)
Correlation	-		-0.066
with Intercept	-		(0.071)
Panel B: School Effects			
Mean	0	0	0.041
	-	-	(0.065)
SD	0.154	0.166	0.105
	(0.005)	(0.006)	(0.008)
Correlation	-		-0.174
with Intercept	-		(0.068)

Notes: The table reports moments of estimated school effects on standardized sixth-grade Math/ELA achievement using Equation (4), pooling the 2014–17 cohorts. Panel A estimates school effects jointly with three co-enrolled-peer FPCA loadings and controls for expected assigned-peer composition; Panel B omits these peer terms. All models include school and cohort effects, school-assignment propensity scores, cubic polynomials in baseline Math and ELA achievement, and student demographics. Column (1) assumes homogeneous effects. In columns (2)–(3), the intercept is the effect for low-achieving students, and the high-achieving-student interaction coefficient is the additional effect for high-achieving students. Intercept means are normalized to zero. Standard deviations and correlations are corrected for fixed-effect estimation error; parentheses report standard errors (delta-method for standard deviations and correlations), and dashes indicate normalized or inapplicable entries.

is consistent with the negative peer effects estimated above. The most effective schools tend to enroll higher-achieving peers, and exposure to higher-achieving peers lowers achievement for high-achieving students. In a school-effect-only model, these negative peer contributions are absorbed into the school-effect estimates. This reduces the dispersion of school-effect estimates for high-achieving students and makes the correlation between the baseline school effect and the

additional effect for high-achieving students more negative.

Moreover, accounting for peer effects leaves school-specific estimates largely unchanged. Appendix Figure A.7 compares school-effect estimates from the peer-adjusted model with those from the school-effect-only model. The two sets of estimates are closely aligned in the homogeneous specification (slope = 0.932, SE = 0.007; R^2 = 0.972) and for low-achieving students in the heterogeneous specification (slope = 0.974, SE = 0.005; R^2 = 0.987). The largest differences appear for high-achieving students. For this group, peer-adjusted school effects tend to exceed school-effect-only estimates among the most effective schools, leading to greater estimated dispersion once peer composition is accounted for. Again, this pattern is consistent with the negative peer effects estimated above: omitting peer composition lowers estimated school effects among schools that enroll higher-achieving peers. Even for this group, however, the two sets of estimates remain closely aligned (slope = 0.886, SE = 0.008; R^2 = 0.958). These results suggest that peer composition is not the main source of dispersion in estimated school effects.

School effects also remain substantial for college enrollment, one of the longer-term outcomes considered in Section 6.1. The estimated standard deviation of peer-adjusted school effects is 6.29 percentage points in the homogeneous model and 6.33 and 8.92 percentage points for high- and low-achieving students, respectively, in the heterogeneous model. These are roughly 8.7-12.4% of the share of college-going students in our data, 72%. School-specific estimates remain strongly correlated with their counterparts from the school-effect-only model; regressions of school-effect-only estimates on peer-adjusted estimates yield slopes of 1.084, 1.178, and 0.995 for the homogeneous, high-achieving, and low-achieving specifications, respectively, with corresponding R^2 values of 0.991, 0.948, and 0.995.[20]

6.3 Alternative Specifications

In this section, we examine the robustness of our results to alternative empirical specifications. We begin by revisiting the modeling choices in Section 2. Our baseline measures peers' achievement relative to the focal student's own achievement. Using absolute peer achievement instead gives us similar peer effect estimates: broadly downward-sloping and larger for high-achieving students

[20]School effects on sixth-grade test scores and college enrollment are positively correlated as well. The correlation is 0.11 between school effects from the homogeneous model and 0.164 for lower-achieving students in the heterogeneous model. The correlation is lower for higher-achieving students at 0.067.

(Appendix Figure A.8). The dispersion of peer-adjusted school effects is also similar across the two specifications: in the homogeneous model, the estimated standard deviation is 0.162σ under both the absolute- and relative-achievement specifications (Appendix Table B.2).

The model also restricts the peer contribution to be a linear functional of the peer-achievement distribution. As discussed in Section 2, this accommodates linear combinations of raw moments but does not nest effects of peer-score variance, a feature examined in prior work (Lyle, 2009). Columns (2) and (3) of Appendix Table B.3 therefore include variance on its own or alongside the FPCA coordinates. Its estimated coefficient is negative in both specifications. But the overall R^2 of the model barely changes, and adding variance to the relative-achievement FPCA model leaves the estimated standard deviation of peer-adjusted school effects nearly unchanged, at 0.163σ compared with 0.162σ in the baseline.

We next examine the FPCA representation introduced in Section 4. Because FPCA selects components to explain variation in peer composition, it may give little weight to features that matter for achievement, including exposure to peers in the tails. We address this concern in two ways. Columns (4)–(7) of Appendix Table B.3 use bottom- and top-tail peer shares at the 5 and 1 percent thresholds, both alone and alongside the FPCA coordinates. We do not find strong evidence that peers at the tail of the distribution matter, and adding these shares leaves the estimated dispersion of peer-adjusted school effects close to the baseline. Appendix Figure A.9 instead approximates $\beta(r)$ directly with cubic B-splines. These estimates preserve the main qualitative patterns, including the broadly downward-sloping function for high-achieving students, although they differ in the tails. Table B.4 further presents results using pre-specified buckets (tercile, quartile) and shows negative peer effects on sixth-grade test scores with the same estimated standard deviation of peer-adjusted school effects.

Within the FPCA framework, we then assess the implementation choices described in Section 5.2. Appendix Figure A.10 varies the number of retained components, the treatment of tail mass, the evaluation grid, and the bandwidth used to estimate peer densities. Across these specifications, the estimated peer-effect function remains similar in shape, and the dispersion of peer-adjusted school effects is nearly unchanged.

Finally, we consider alternative subjects and student groups. Estimating the model separately

for Math and ELA yields the broadly negative peer-effect pattern found for the combined achievement measure (Appendix Figure A.11). Extending the heterogeneity analysis to sex and race and ethnicity reveals modest, imprecisely estimated differences by sex (Appendix Figure A.12). Across racial and ethnic groups, the estimated function for Asian students is broadly downward sloping, while that for Hispanic students is nonmonotonic and imprecisely estimated.

7. Relative Contributions of Schools and Peers to Achievement

With the estimates in hand, we next ask the following two questions. First, how much achievement gain can an *individual* family expect through the school and peer channels, respectively, by sending their kid to different schools? Furthermore, how much *systemwide* achievement gain can we obtain when we potentially reassign every student to schools and different peer groups?

7.1 From an Individual Family's Perspective

First, to assess the potential gains available to an individual family, we consider reassigning one student at a time to each school in our sample, holding all other students' assignments fixed. Each reassignment changes both the school's contribution to the student's achievement and the peer group the student is exposed to. We assess the relative scope for gains through the school and peer channels by comparing these two components.

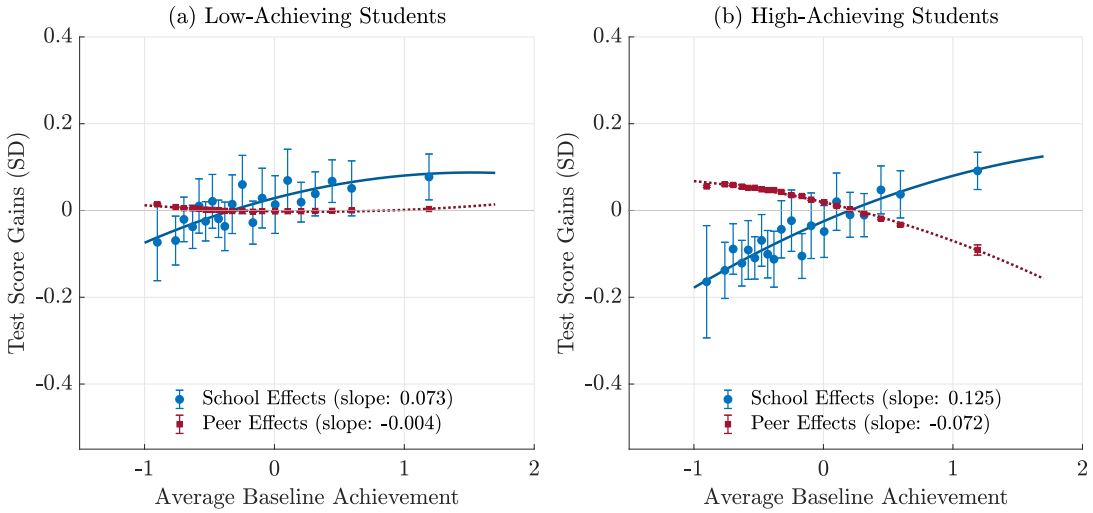
We randomly sample ten students from each percentile of baseline achievement within each cohort. For each sampled student and destination school, we calculate the contributions of the destination school's peer-adjusted effect and incumbent peer composition to the student's predicted achievement, using empirical-Bayes shrunken school effects and the peer-effect function from our heterogeneous relative-achievement model. We express the contribution through each channel as a gain relative to the student's observed assignment. Separately for low- and high-achieving students, we then average these gains across sampled movers for each destination school and compare their dispersion across destinations, weighting schools equally.

The results show that peer-adjusted school effects offer substantially greater scope for gains across destinations than peer effects. For low-achieving students, the standard deviation of destination-specific average peer-effect gains is about 3.4% of the corresponding standard deviation of school-effect gains. The ratio is larger for high-achieving students, at about 23%, but

gains through the school channel remain substantially more dispersed.

Notably, families generally observe neither school nor peer contributions when deciding where to apply. We therefore ask how predicted gains through the two channels vary with information families often use to guide their choices—the average baseline achievement of enrolled students (Abdulkadiroğlu et al., 2020; Ainsworth et al., 2023; Campos, 2024).

Figure 4: One-Student Reassignment Gains by Average Baseline Achievement



Notes: Panels (a) and (b) report predicted gains for low- and high-achieving students, respectively, using the heterogeneous relative-achievement model. We sample ten students per baseline-achievement percentile and cohort and reassign each to every destination school, holding other students' assignments fixed. Gains are measured relative to each mover's observed assignment, separately through empirical-Bayes peer-adjusted school effects (blue) and peer composition (burgundy), in achievement standard deviations. We first average each gain component across sampled movers for each destination. Markers then average these school-level gains within 20 common quantile bins of destination schools' average baseline achievement, weighting schools equally. Vertical bars are pointwise 95% confidence intervals for each bin mean, computed from variation across schools within the bin. These intervals do not incorporate uncertainty from estimating the school and peer effects or preceding estimation stages. The solid blue and dotted burgundy curves are quadratic fits to the unbinned school averages; legend slopes come from separate linear regressions on those averages.

Figure 4 presents the average gain from each channel across schools in each ventile bin, and its associated 95% confidence interval. For low-achieving students, Panel (a) shows that peer-effect gains remain close to zero across schools with different average baseline achievement. Average school-effect gains generally increase with baseline achievement and also vary substantially across schools within the same achievement bin.

Panel (b) shows that schools with higher average baseline achievement tend to offer larger positive school-effect gains but more negative peer-effect gains. This has two implications. First,

families using average achievement as a proxy for a school's contribution to learning therefore face offsetting changes in the two components. Second, the most selective schools need not maximize total predicted gains.[21] However, overall, school-effect gains vary more than peer-effect gains both across achievement bins and across schools within the same bin even for this group.

7.2 Optimal Assignment

We now consider the problem of optimally reassigning students across schools to maximize aggregate achievement. While flexible peer effects are essential for us to expect any aggregate gain under an alternative peer assignment, they also make the optimal assignment problem generally intractable. Each reassignment changes the moving student's school and peers while also affecting classmates at both the origin and destination schools. The aggregate gain from a move therefore depends on which other students move alongside it: a reassignment that improves achievement under one peer configuration may reduce it under another. The planner must jointly determine student–school matches and the peer groups those matches create, across an exponentially large set of feasible assignments.

We use a low-rank continuum approximation to obtain a computationally manageable formulation. First, rather than choosing individual student assignments directly, the planner chooses a baseline-achievement density for each school. Second, we apply FPCA to the observed school-level baseline-achievement densities and represent each candidate density by its loadings on the first three components. This replaces the individual-level combinatorial assignment problem with a continuous choice over three school-composition coordinates per school.

FPCA plays a second role beyond dimension reduction by focusing both estimation and policy analysis on empirically informative variation. Rather than attempting to estimate achievement responses to arbitrary perturbations of the peer distribution, our peer-effect model concentrates on the leading components that account for most of the observed variation. The assignment exercise similarly allows school compositions to change along the leading components of observed cross-school variation. Together with the convex-hull restriction imposed below, this creates a disciplined, data-adaptive policy space in which both the response estimates and the counterfactual

[21] This is consistent with Abdulkadiroğlu et al. (2014) that shows marginal students do not experience improvement in their achievement in the most selective schools in similar settings.

changes used to compute assignment gains are well represented in the data.

Given these school-composition vectors, the school and peer components of predicted achievement have a closed-form quadratic representation. The school-match term is linear in the loadings, while the peer-effects term is quadratic because each school's density determines both the recipients and their peers. Pooling the three cohorts, the resulting problem has roughly 1,500 choice variables,[22] with feasibility restrictions entering as linear constraints. Under our estimates, the resulting problem is concave, so the computed optimum is the global optimum of this continuous problem (See Appendix H for details).

Using this idea, we can define the planner's problem formally as follows. For each school j, the planner chooses a vector of absolute-FPCA loadings, $\mathbf{a}_j = (a_{1j}, \ldots, a_{Mj})$. These loadings determine a counterfactual baseline-achievement density $h_j(\cdot; \mathbf{a}_j)$, given by

$$h_j(x; \mathbf{a}_j) = \bar{h}(x) + \sum_{m=1}^{M} a_{mj} v_m^a(x),$$

for each x in the support of baseline achievement. Here, $\bar{h}$ is the citywide distribution of baseline achievement, and $\{v_m^a\}_{m=1}^M$ are the FPCA basis functions constructed from the corresponding centered school-level distributions. We use the superscript a to distinguish these basis functions from the relative-achievement basis functions used in estimation. Note that every candidate $h_j(x; \mathbf{a}_j)$ integrates to one regardless of the chosen loadings, because $\bar{h}$ integrates to one and the basis functions integrate to zero.

We evaluate each counterfactual composition using the estimates from our heterogeneous relative-score model. The relevant inputs are the estimated peer-effect functions, $\widehat{\beta}^L(r)$ and $\widehat{\beta}^H(r)$, and the estimated peer-adjusted school effects, $\widehat{\mu}_j^L$ and $\widehat{\mu}_j^H$, for low- and high-achieving students.

By pinning down the distribution of baseline achievement among students assigned to school j, the loadings $\mathbf{a}_j$ determine both the school-effect contribution and the peer-effect contribution of that school's composition. First, the loadings determine the share of low-achieving students assigned to school j, $L_j(\mathbf{a}_j)$, and hence the school-effect contribution $L_j(\mathbf{a}_j)\widehat{\mu}_j^L + \big(1 - L_j(\mathbf{a}_j)\big)\widehat{\mu}_j^H$. Second, the loadings determine the distribution of within-school relative achievement gaps, $r = x' - x$, by

[22] This is the number of schools (502) times the number of FPCs (3)

determining the density of recipient scores x and peer scores x' within school j. The average peer-effect contribution at school j is

$$P_j(\mathbf{a}_j) = \int\int \left[\mathbf{1}\{x < x_{\text{med}}\}\widehat{\beta}^L(x' - x) + \mathbf{1}\{x \geq x_{\text{med}}\}\widehat{\beta}^H(x' - x)\right] h_j(x; \mathbf{a}_j) h_j(x'; \mathbf{a}_j)\, dx'\, dx.$$

This expression evaluates peer effects using the full school-level density $h_j(\cdot; \mathbf{a}_j)$ for both recipients and peers. In estimation, peer exposure is measured using leave-one-out peer distributions. In this counterfactual exercise, we abstract from that finite-sample correction and approximate each student's peer distribution by the school-level composition.[23]

Combining the peer effect and peer-adjusted school effects components, aggregate predicted achievement given a vector of school loadings $(\mathbf{a}_1, \ldots, \mathbf{a}_J)$ is

$$W(\mathbf{a}_1, \ldots, \mathbf{a}_J) = \sum_{j=1}^{J} K_j \left[L_j(\mathbf{a}_j)\widehat{\mu}_j^L + \big(1 - L_j(\mathbf{a}_j)\big)\widehat{\mu}_j^H + P_j(\mathbf{a}_j)\right],$$

where K_j is the fixed capacity[24] of school j. The planner's problem can therefore be written as

$$\begin{aligned} \max_{\mathbf{a}_1, \ldots, \mathbf{a}_J} \quad & W(\mathbf{a}_1, \ldots, \mathbf{a}_J) \\ \text{s.t.} \quad & \sum_{j=1}^{J} K_j \mathbf{a}_j = \sum_{j=1}^{J} K_j \mathbf{a}_j^{\text{obs}}, \\ & \mathbf{a}_j \in \text{conv}\{\mathbf{a}_1^{\text{obs}}, \ldots, \mathbf{a}_J^{\text{obs}}\}, \qquad j = 1, \ldots, J. \end{aligned} \tag{5}$$

Here, $\mathbf{a}_j^{\text{obs}}$ denotes the vector of FPCA loadings corresponding to the observed distribution of baseline achievement at school j, and $\text{conv}\{\mathbf{a}_1^{\text{obs}}, \ldots, \mathbf{a}_J^{\text{obs}}\}$ denotes the convex hull of these school-specific loading vectors. Because each basis function integrates to zero, school j's enrollment remains fixed at K_j regardless of $\mathbf{a}_j$.

The first constraint preserves the citywide distribution of baseline achievement, as represented

[23] This avoids making the objective depend on individual-level assignments within each school; the omitted leave-one-out correction is of order one over the size of a school's cohort.

[24] We use observed enrollment as the relevant capacity constraint, so we compare student outcomes across assignments in which each school has the same number of enrolled students.

by the retained FPCA components, across alternative school-level peer compositions.[25] The second constraint restricts each candidate peer distribution to be a convex combination of the J school-level distributions observed in the data, under the retained FPCA representation.[26] This restriction keeps counterfactual peer distributions close to the empirical distributions used to estimate the model, rather than allowing arbitrary peer distributions for which the estimated peer effects would require substantially stronger extrapolation. It also preserves nonnegativity of the candidate peer distributions on the achievement grid, provided the FPCA reconstructions of the observed school-level distributions are nonnegative on that grid. Appendix Section H provides additional details on the numerical implementation.

To benchmark the solution to Equation (5), we also solve auxiliary optimization problems that isolate each component of predicted achievement. Holding the feasible set of loadings fixed, we separately maximize and minimize the peer-adjusted school-effect component and the peer-effect component. We then characterize the tradeoff between the two channels: For each peer-effect value, we select the feasible assignment that maximizes the school-effect component; conversely, for each school-effect value, we select the feasible assignment that maximizes the peer-effect component. These help trace the frontier of feasible combinations of peer and school value. We report both components as achievement gains relative to their values under the observed assignment.

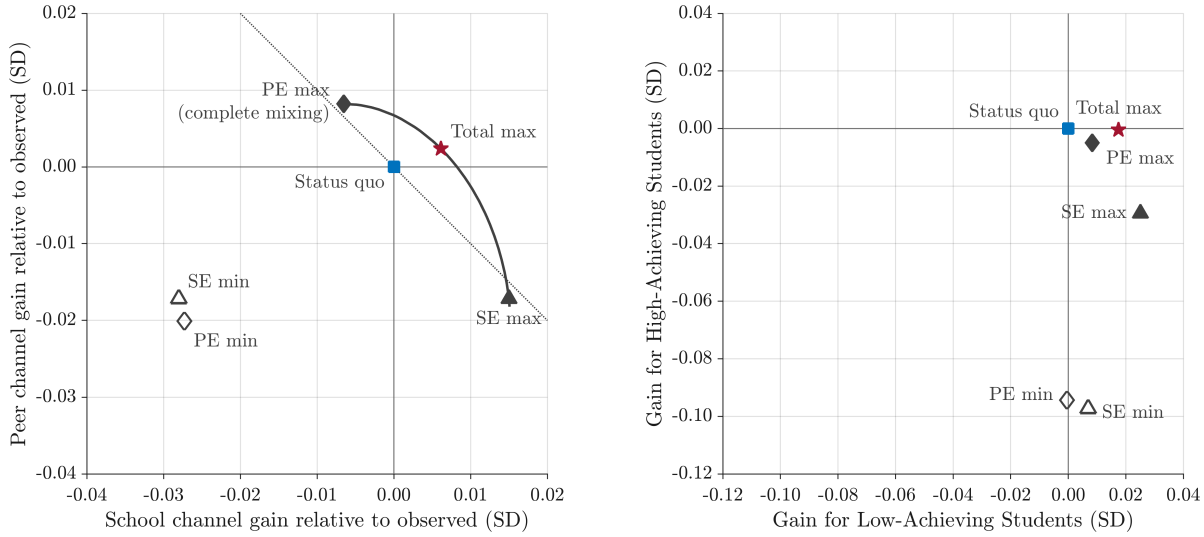
Panel (a) of Figure 5 plots combinations of peer-effect and peer-adjusted school-effect contributions, normalized relative to the observed assignment. The achievement-maximizing allocation increases predicted achievement by about 0.0085 SD per student relative to the observed assignment. Most of this gain comes through the school-effect channel, which rises by 0.0061 SD, while the peer channel rises by 0.0024 SD. Because our school-effect model allows effects to differ by baseline achievement, the school-effect component rewards assigning low- and high-achieving students to the schools with the largest estimated effects for their group. The peer-effect component instead favors more mixed peer compositions. At the optimum both channels improve, and student–school matching is the main source of gains, consistent with the larger dispersion in

[25] The aggregate density at each point is $\sum_j K_j h_j(x; \mathbf{a}_j) = \bar{h}(x)\left(\sum_j K_j\right) + \sum_m v_m^a(x)\left(\sum_j K_j a_{mj}\right)$. Thus the constraint $\sum_j K_j \mathbf{a}_j = \sum_j K_j \mathbf{a}_j^{\text{obs}}$ keeps fixed the aggregate FPCA component of the citywide baseline-achievement distribution.

[26] Because $h(x; \mathbf{a})$ is affine in $\mathbf{a}$, any $\mathbf{a} \in \text{conv}\{\mathbf{a}_1^{\text{obs}}, \ldots, \mathbf{a}_J^{\text{obs}}\}$ implies $h(x; \mathbf{a}) = \sum_{j=1}^J \lambda_j h(x; \mathbf{a}_j^{\text{obs}})$ for some weights $\lambda_j \geq 0$ with $\sum_{j=1}^J \lambda_j = 1$.

peer-adjusted school effects. Away from the optimum, the two channels trade off: moving further toward the school-effect maximum lowers the peer contribution sharply.

Figure 5: Achievement Gains from Feasible Student Reassignments



(a) Gains by Channel

(b) Gains by Achievement Group

Notes: All gains are relative to the observed assignment and measured in achievement standard deviations. In Panel (a), the horizontal and vertical axes report gains through peer-adjusted school effects and peer effects, respectively, per student in the full sample. The solid curve traces the boundary of feasible combinations under fixed school capacities, a fixed aggregate FPCA representation of baseline achievement, and school compositions restricted to the convex hull of observed compositions. Panel (b) reports total gains, combining school and peer effects, for low-achieving students on the horizontal axis and high-achieving students on the vertical axis, per student within each group. Weighting these group-specific gains by their sample shares gives the overall gain, equal to the sum of the two coordinates in Panel (a). In both panels, the square marks the observed assignment, the star the allocation that maximizes total achievement, diamonds the peer-component extrema, and triangles the school-component extrema. Appendix H describes how the extrema are computed . In Panel (a), the peer-component maximum coincides with complete mixing, in which every school has the citywide baseline-achievement distribution, and the dotted diagonal marks zero total gain relative to the observed assignment.

To quantify the gains that peer redistribution can generate on its own, we consider the allocation that maximizes the peer-effect component. This allocation coincides with complete mixing, in which every school has the citywide baseline-achievement distribution. It raises the peer component by about 0.008 SD per student relative to the observed assignment, but lowers the school component by about 0.007 SD. For comparison, the lowest peer component we find is about 0.020 SD per student below the observed assignment. Thus, within the feasible set, the peer-effect component ranges by about 0.028 SD per student. The observed assignment is already relatively close

to the peer-effect maximum, capturing about 71% of the feasible peer-effect range.

The corresponding range is about 1.5 times as large for the peer-adjusted school-effect component. Maximizing the school component raises it by about 0.015 SD per student relative to the observed assignment, but lowers the peer component by 0.017 SD, for a net loss of 0.002 SD. For comparison, the minimum of the school component is about 0.028 SD per student below the observed assignment. Thus, within the feasible set, the school-effect component ranges by about 0.043 SD per student. The observed assignment captures about 65% of this range. Although the observed assignment captures a large share of both feasible ranges, the school-effect range is larger, leaving more remaining upside in absolute terms through the school-effect channel than through the peer-effect channel.

Panel (b) of Figure 5 shows how these gains are distributed across baseline-achievement groups. At the allocation that maximizes total achievement, the gains accrue almost entirely to low-achieving students, whose predicted achievement rises by about 0.017 SD, while achievement for high-achieving students is essentially unchanged. This is because while both groups face substantial dispersion in peer-adjusted school effects, peer effects are muted for low-achieving students and predominantly negative for high-achieving students. At the total-achievement maximum, low-achieving students are assigned to more effective schools and the resulting school effectiveness loss for displaced high-achieving students is largely offset by being mixed with low-achieving peers who increase their test scores. Other assignments reveal a sharper distributional tradeoff. Maximizing the school-effect component raises achievement for low-achieving students by about 0.025 SD but lowers it for high-achieving students by about 0.029 SD; maximizing the peer-effect component produces corresponding gains and losses of approximately 0.008 and 0.005 SD. Thus, the modest aggregate gains from reassignment can mask substantial differences across achievement groups, and maximizing either component in isolation need not improve outcomes for both groups.

8. Conclusion

Many educational policies change both the school a student attends and the peers with whom she learns. This paper proposes a framework that combines lottery-induced variation in school and

assigned-peer composition with a functional principal component representation of peers' prior-achievement distributions. This framework can in principle exploit quasi-random tie-breakers, and we apply it to the New York City middle school assignment system.

We find that exposure to higher-achieving peers does not uniformly improve sixth-grade achievement, and the response varies with students' own baseline achievement. Yet accounting for peers changes the dispersion and ranking of estimated school effects relatively little. In this setting, the schools that appear effective do not owe most of that advantage to differences in their students' prior achievement.

This distinction between school productivity and peers matters for an individual family's school choice as well as for optimal assignment. Attending a school with higher-achieving classmates and attending a school that raises achievement are not the same, and if anything, families face a trade-off between the two in this setting. Systemwide, the best computed allocation raises predicted achievement by about 0.009 standard deviations per student, about 70 percent of which comes from matching students to schools with larger peer-adjusted effects for their prior-achievement group. Importantly, maximizing one dimension can lower gains through the other, highlighting the importance of the unified perspective.

References

Abdulkadiroğlu, Atila, Joshua Angrist, and Parag Pathak, "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," *Econometrica*, 2014, *82* (1), 137–196.

_ , Joshua D Angrist, Yusuke Narita, and Parag A Pathak, "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation," *Econometrica*, 2017, *85* (5), 1373–1432.

Abdulkadiroglu, Atila, Parag A Pathak, and Christopher R Walters, "Who Gets What in Education: Can School Matching Improve Student Achievement?," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2026, (2026-38).

Abdulkadiroğlu, Atila, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters, "Do Parents Value School Effectiveness?," *American Economic Review*, 2020, *110* (5), 1502–1539.

Ainsworth, Robert, Rajeev Dehejia, Cristian Pop-Eleches, and Miguel Urquiola, "Why Do Households Leave School Value Added on the Table? The Roles of Information and Preferences," *American Economic Review*, 2023, *113* (4), 1049–1082.

Allende, Claudia, "Competition Under Social Interactions and the Design of Education Policies," 2019,

p. 104.

Angrist, Joshua D., "The Perils of Peer Effects," *Labour Economics*, October 2014, *30*, 98–108.

_ , Peter D. Hull, Parag A. Pathak, and Christopher R. Walters, "Leveraging Lotteries for School Value-Added: Testing and Estimation*," *The Quarterly Journal of Economics*, May 2017, *132* (2), 871–919.

Angrist, Joshua, Peter Hull, and Christopher Walters, "Methods for Measuring School Effectiveness," in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Handbook of the Economics of Education*, Vol. 7, Elsevier, 2023, chapter 1, pp. 1–60.

_ , _ , Parag A. Pathak, and Christopher Walters, "Credible School Value-Added with Undersubscribed School Lotteries," *Review of Economics and Statistics*, January 2024, *106* (1), 1–19.

Bastos, Paulo, Julian Cristia, Beomsoo Kim, and Ofer Malamud, "Good schools or good students? Evidence on school effects from universal random assignment of students to high schools," Technical Report, IDB Working Paper Series 2022.

Bibler, Andrew, Stephen B. Billings, and Stephen L. Ross, "When Opportunity Costs Others: The Spillover Effects of School Choice Lotteries on Criminality," 2026. Manuscript, March 11, 2026.

Booij, Adam S., Edwin Leuven, and Hessel Oosterbeek, "Ability Peer Effects in University: Evidence from a Randomized Experiment," *The Review of Economic Studies*, 2017, *84* (2), 547–578.

Borusyak, Kirill and Peter Hull, "Nonrandom Exposure to Exogenous Shocks," *Econometrica*, 2023, *91* (6), 2155–2185.

Boucher, Vincent, Michelle Rendall, Philip Ushchev, and Yves Zenou, "Toward a general theory of peer effects," *Econometrica*, 2024, *92* (2), 543–565.

Burke, Mary A. and Tim R. Sass, "Classroom Peer Effects and Student Achievement," *Journal of Labor Economics*, 2013, *31* (1), 51–82.

Campos, Christopher, "Social Interactions, Information, and Preferences for Schools: Experimental Evidence from Los Angeles," Working Paper 33010, National Bureau of Economic Research 2024. Revised November 2025.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West, "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 2013, *81* (3), 855–882.

Chang, Yoosoon, Steven N Durlauf, Seunghee Lee, and Joon Y Park, "A Trajectories-Based Approach to Measuring Intergenerational Mobility," Technical Report, National Bureau of Economic Research 2023.

Epple, Dennis and Richard E. Romano, "Chapter 20 - Peer Effects in Education: A Survey of the Theory and Evidence," in Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, eds., *Handbook of Social Economics*, Vol. 1, North-Holland, January 2011, pp. 1053–1163.

Graham, Bryan S., Guido W. Imbens, and Geert Ridder, "Measuring the Effects of Segregation in the Presence

of Social Spillovers: A Nonparametric Approach," *Journal of Econometrics*, 2025, *252*, 106006.

Hoxby, Caroline and Gretchen Weingarth Salyer, "School Reassignment and the Structure of Peer Effects, Manuscrito," *Universidad de Harvard*, 2006.

Imberman, Scott A., Adriana D. Kugler, and Bruce I. Sacerdote, "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees," *American Economic Review*, 2012, *102* (5), 2048–2082.

Jackson, C. Kirabo, "Can Higher-Achieving Peers Explain the Benefits to Attending Selective Schools? Evidence from Trinidad and Tobago," *Journal of Public Economics*, 2013, *108*, 63–77.

Lavy, Victor, M. Daniele Paserman, and Analia Schlosser, "Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom," *The Economic Journal*, 2012, *122* (559), 208–237.

Lazear, Edward P, "Educational production," *The quarterly journal of economics*, 2001, *116* (3), 777–803.

Lyle, David S., "The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point," *American Economic Journal: Applied Economics*, October 2009, *1* (4), 69–84.

Sacerdote, Bruce, "Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?," in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Handbook of the Economics of Education*, Vol. 3, Elsevier, January 2011, pp. 249–277.

Sargan, J. D., "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 1958, *26* (3), 393–415.

Tincani, Michela M., "Peer Effects and Rank Concerns in the Classroom," *Journal of Political Economy*, 2026. Advance online publication.

Walters, Christopher, "Empirical Bayes methods in labor economics," in "Handbook of Labor Economics," Vol. 5, Elsevier, 2024, pp. 183–260.

Woodcock, Simon, Jane Friesen, and Ricardo Meilman Lomaz Cohn, "Sorting, Peer Effects and School Effectiveness in Private and Public Schools," May 2025.

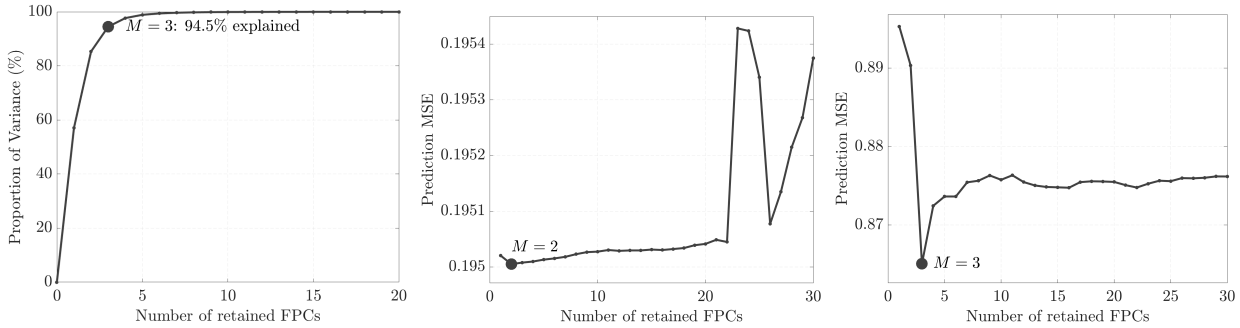
ONLINE APPENDIX

Disentangling School and Peer Effects: A Non-Linear Approach with School Lotteries

Mariana Laverde, Minseon Park, & Suk Joon Son

A. Additional Figures

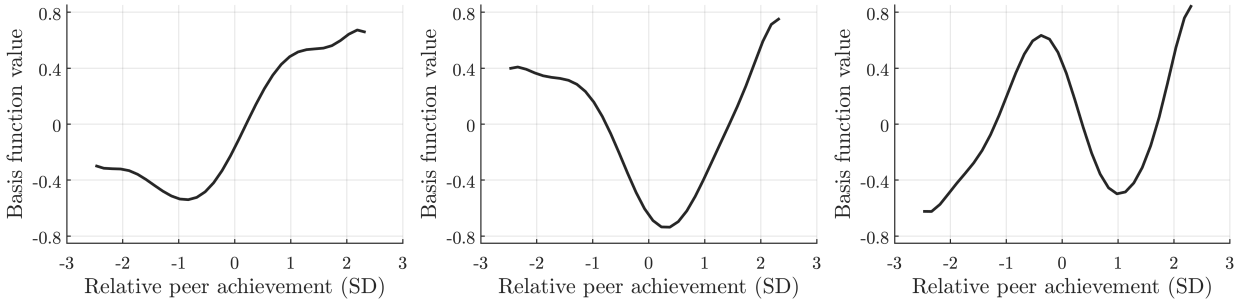
Figure A.1: Choosing the Number of Retained FPCA Components for Enrolled Peer Densities



(a) Variance explained (b) Outcome-prediction MSE (c) Structural-prediction MSE

Notes: The figure reports diagnostics used to choose the number of enrolled-peer-density FPCA components in the pooled 2014–2017 relative-score MathELA specification. Panel (a) plots cumulative variation explained by the leading components and marks the default choice, $M = 3$, which explains 94.5% of the variation. Panels (b) and (c) report out-of-fold prediction mean squared error from 10-fold cross-validation. Panel (b) uses the controlled outcome-prediction exercise, with per-fold residualization. Panel (c) uses the structural-prediction exercise based on the identified peer-effect coefficients, controlled school fixed effects, and fold-level intercept correction. The marker in each cross-validation panel denotes the MSE-minimizing number of components. Enrolled-peer densities retain the central 95 percent of the mean density, with score values below 2.5 percent and above 97.5 percent trimmed before FPCA. Error bars are omitted in Panels (b) and (c).

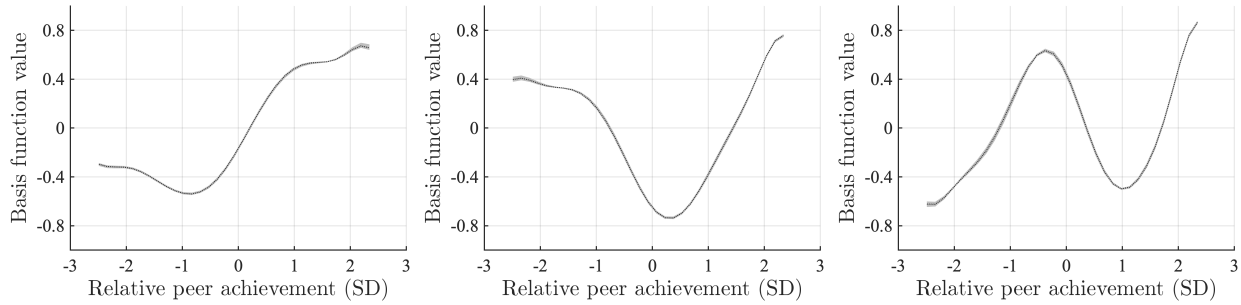
Figure A.2: Three Leading Functional Principal Components of Peer Distributions



(a) First component $v_1(r)$ (b) Second component $v_2(r)$ (c) Third component $v_3(r)$

Notes: This figure plots the first three functional principal components estimated from the peer-achievement distributions. Panels (a), (b), and (c) show $v_1(r)$, $v_2(r)$, and $v_3(r)$, respectively, where r is relative peer achievement measured in standard deviation units. Components are ordered by the share of variation in peer distributions they explain. The sign of each component is arbitrary and should be interpreted relative to the corresponding loading.

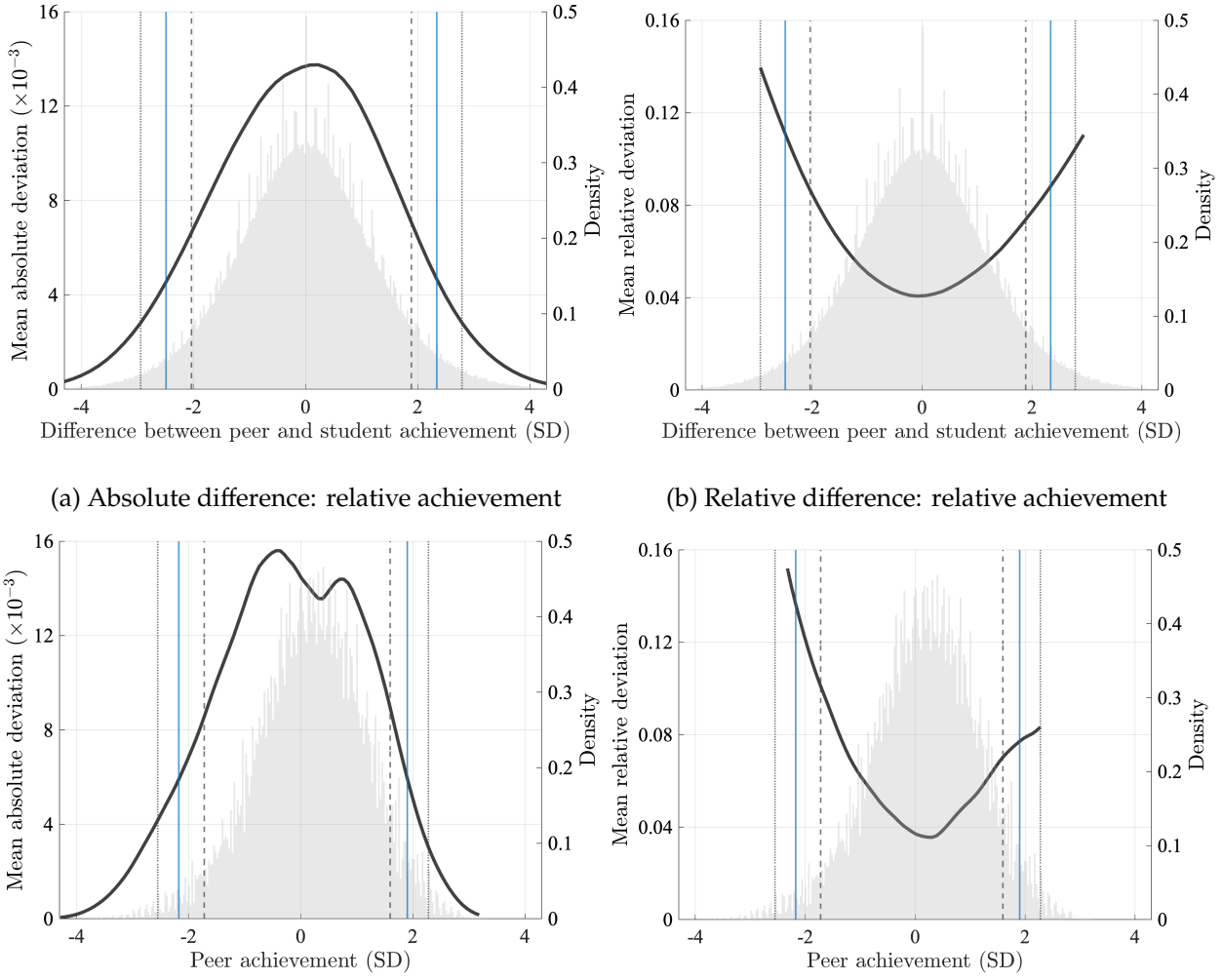
Figure A.3: Bootstrap Stability of the Three Leading Functional Principal Components



(a) First component $v_1(r)$ (b) Second component $v_2(r)$ (c) Third component $v_3(r)$

Notes: This figure compares the first three functional principal components of co-enrolled peer distributions across 100 school-cluster bootstrap samples in the baseline relative-achievement specification. Thin dark dotted curves show the full-sample estimates and gray curves show the bootstrap estimates. Because FPCs remain stable across bootstrap samples, gray curves from the bootstrap estimates stay close to the original FPCs. Schools are resampled with replacement, holding the student-level peer-density inputs and support grid fixed. The sign of each bootstrap component is aligned with its full-sample counterpart.

Figure A.4: Pointwise Difference Between Assigned and Expected Assigned-Peer Densities



(a) Absolute difference: relative achievement

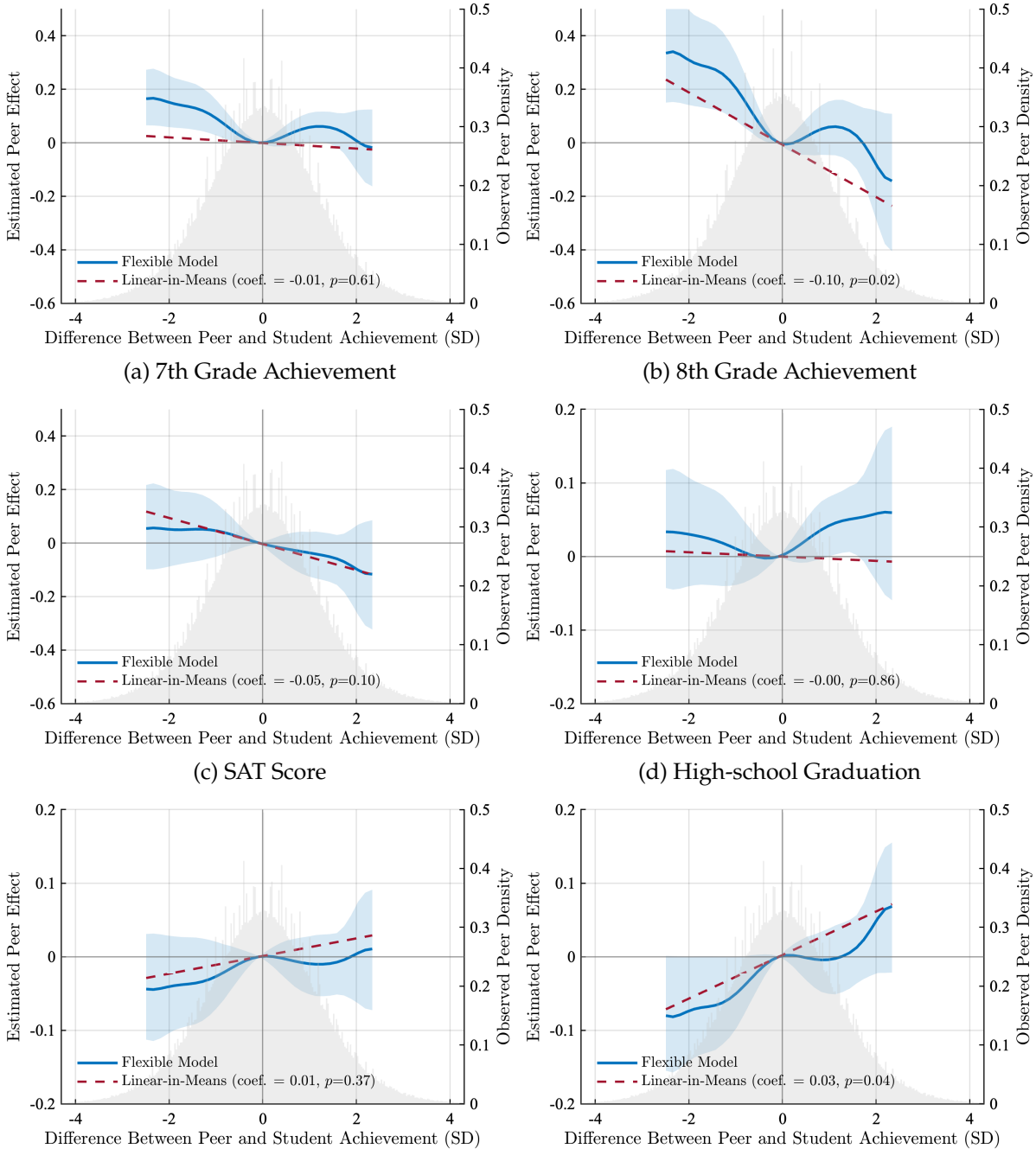
(b) Relative difference: relative achievement

(c) Absolute difference: absolute achievement

(d) Relative difference: absolute achievement

Notes: The figure summarizes quasi-random variation in assigned peer distributions. The top row plots, at each score-grid point, the cross-student mean absolute difference between realized assigned peer densities and expected assigned peer densities. This difference measures the amount of assignment-induced variation in peer exposure at each point of the distribution. The bottom row reports the corresponding relative difference, obtained by scaling the absolute difference by the expected co-assigned peer density at the same grid point. Panels (a) and (b) use peer achievement relative to the focal student, while Panels (c) and (d) use absolute peer achievement. Each density is normalized to integrate to one before computing these differences. The gray histogram shows the corresponding distribution of realized co-assigned peer achievement, with density reported on the right axis. In the top row, values on the left axis are reported in units of 10^{-3}. Paired dotted, solid, and dashed vertical lines mark the 1, 2.5, and 5 percent tail cutoffs, respectively, of the mean enrolled-peer density. In the baseline estimation, we restrict co-enrolled peer densities to the interval between the 2.5th and 97.5th percentiles of the pooled enrolled-peer density, marked by the solid vertical lines, and redistribute each student's omitted tail mass over the retained support using boundary-concentrated normal weights.

Figure A.5: Peer Effects on Later Achievement and Educational Attainment



(a) 7th Grade Achievement

(b) 8th Grade Achievement

(c) SAT Score

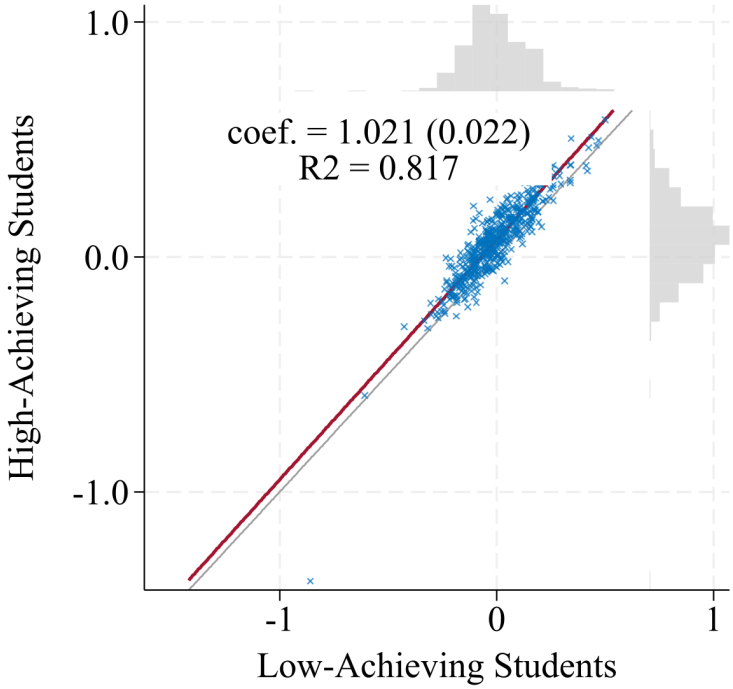
(d) High-school Graduation

(e) Out-of-state college enrollment

(f) New York private-college enrollment

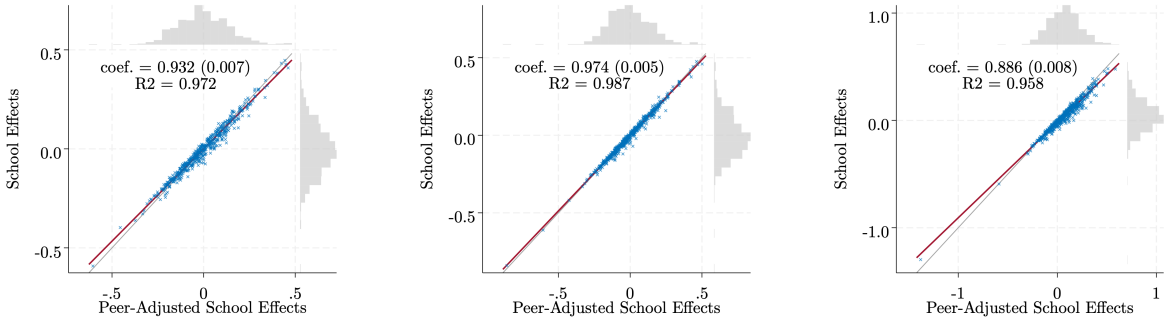
Notes: Each panel reports estimated peer effects for the outcome named below. The blue curve is the flexible FPCA estimate using three enrolled-density components, the shaded area is its pointwise 95 percent confidence interval, and the burgundy dashed line is the linear-in-means estimate. The gray histogram shows the observed distribution of relative peer achievement. Outcomes in Panels (a)–(c) are standardized; the binary outcomes in Panels (d)–(f) remain in 0–1 units. All regressions include expected-density FPCA coordinates, school-assignment propensity scores, year and school fixed effects, cubic baseline math and ELA scores, and student demographic controls.

Figure A.6: Peer-Adjusted School Effects for High- and Low-Achieving Students



Notes: Each marker is a school. The horizontal axis reports the empirical-Bayes shrunken school effect for low-achieving students, and the vertical axis reports the corresponding effect for high-achieving students, both from the relative-achievement school-plus-peer model. The two components are estimated jointly; the high-achieving-student effect is the sum of the shrunken baseline and achievement-group interaction components. Shrinkage uses the bias-corrected between-school covariance and each school's fixed-effect covariance matrix. The gray line is the 45-degree line, the maroon line is an equally weighted OLS fit across schools, and marginal histograms show the two school-effect distributions. The box reports the fitted slope, its standard error, and R^2.

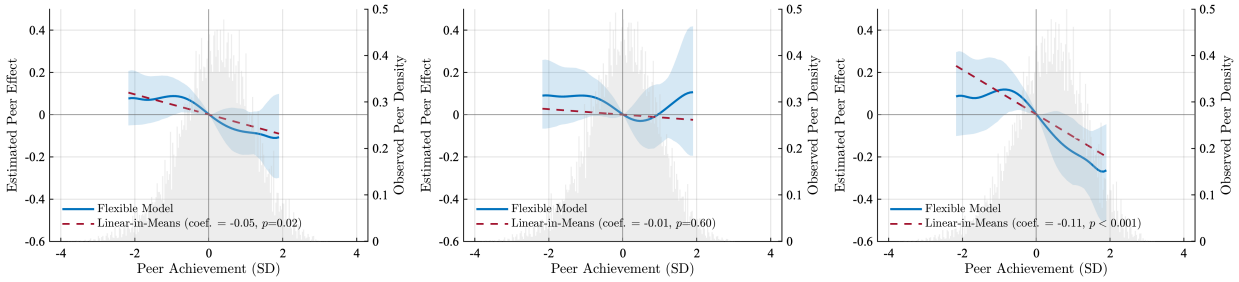
Figure A.7: Comparison of Peer-Adjusted and School-Only School Effect Estimates



(a) Full Sample (b) Low-Achieving Students (c) High-Achieving Students

Notes: Each marker is a school. The horizontal axis reports its empirical-Bayes-shrunken school effect from the peer-adjusted model; the vertical axis reports its shrunken effect from the school-only model. Panels (a)–(c) use the full sample, low-achieving students, and high-achieving students, respectively. The gray line is the 45-degree line, the maroon line is an equally weighted OLS fit across schools, and the marginal histograms show the two school-effect distributions. The inset reports the fitted slope, its standard error, and R^2.

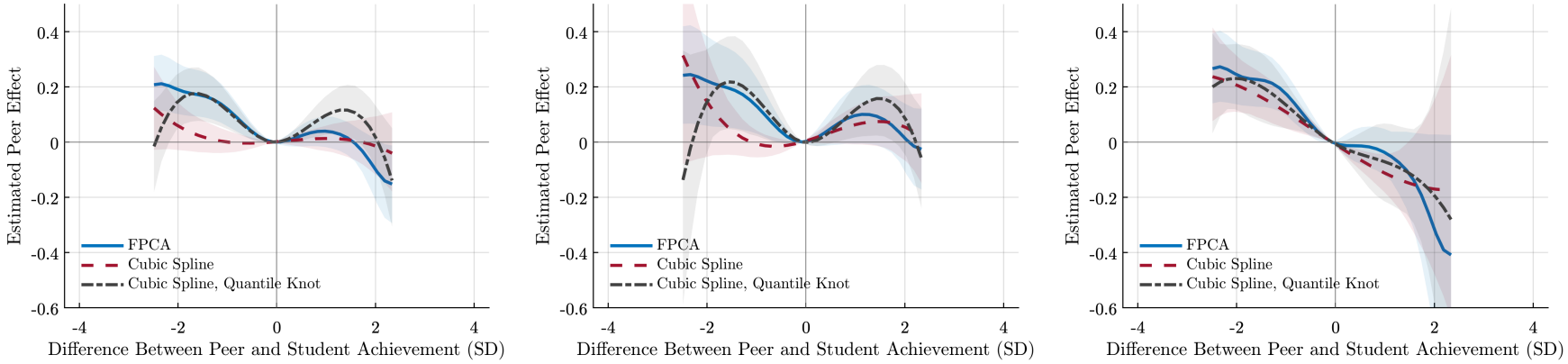
Figure A.8: Peer Effects: Absolute Peer Achievement



(a) Full sample (b) Low-Achieving Students (c) High-Achieving Students

Notes: Panels (a)–(c) report estimates for the full sample, low-achieving students, and high-achieving students. Peer achievement is measured in absolute standardized prior-achievement units rather than relative to the focal student's prior achievement. The solid blue curve is the three-component FPCA estimate, the dashed burgundy line is the linear-in-means estimate, the blue band is a pointwise 95 percent confidence interval, and the gray histogram is the observed distribution of co-enrolled peers' baseline achievement. Both functions are normalized to zero at peer achievement zero. The models otherwise follow the main peer-effect specifications.

Figure A.9: Alternative Methods for Summarizing Peer Effects



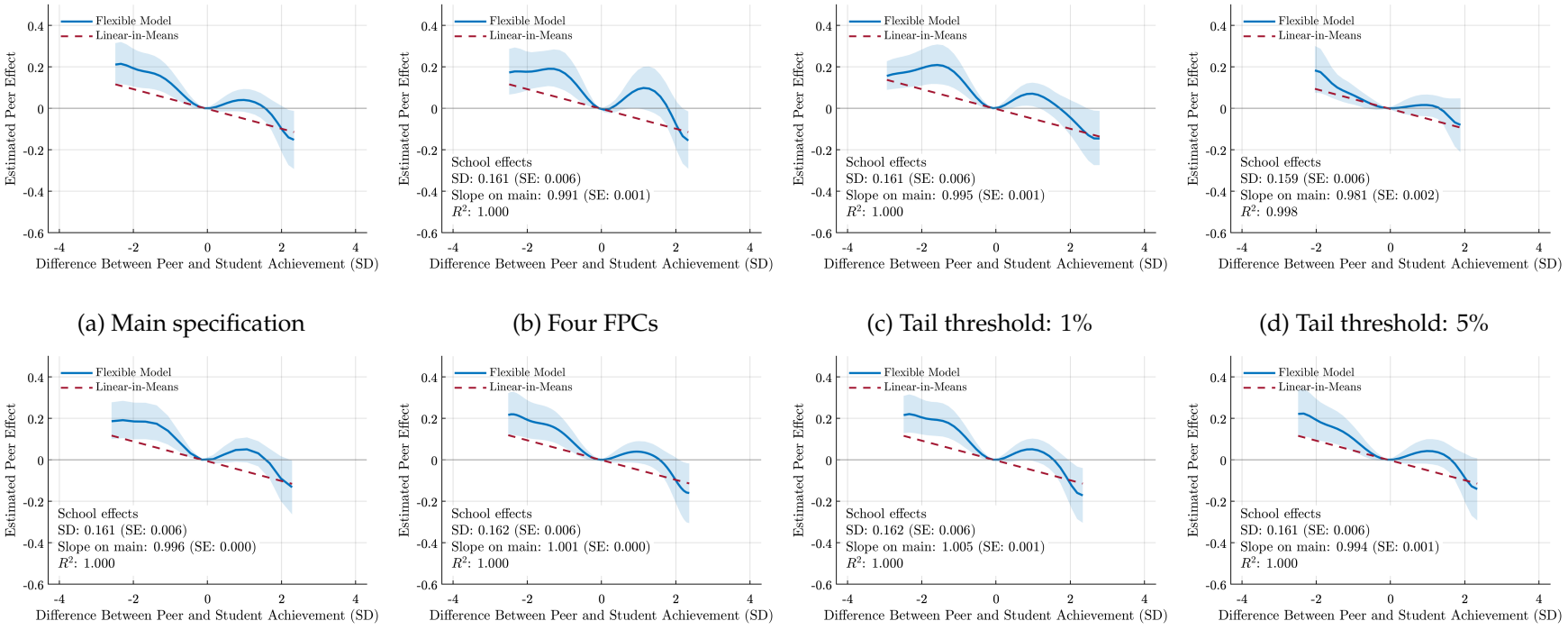
(a) Full sample (b) Lower-Achieving Students (c) Higher-Achieving Students

Notes: This figure compares the baseline three-component FPCA specification with two cubic B-spline specifications for the peer-effect function $\beta(r)$, where r is peer achievement relative to the student's own achievement. The first spline has no interior knots and therefore reduces to a single cubic polynomial. The second has one interior knot at the median of the pooled peer-score density, allowing two cubic pieces joined with continuous first and second derivatives.

For a spline approximation $\beta(r) \approx \sum_k b_k B_k(r)$, the FPCA scores are replaced by $z_{ik} = \int B_k(r) f_i(r)\, dr$, where f_i is student i's relative peer-score density. Because the basis functions sum to one and each density integrates to one, one redundant regressor is omitted. The two spline specifications therefore have three and four free coefficients, respectively. Expected peer densities are summarized using the corresponding spline basis. All specifications include school and year fixed effects, student controls, propensity scores, and expected peer-distribution controls.

Panel (a) reports pooled estimates for the 2014–17 MathELA sample; panels (b) and (c) report estimates for lower- and higher-achieving students, split at median baseline achievement. Each curve displays $\hat{\beta}(r) - \hat{\beta}(0)$, normalizing the effect of peers with the student's own achievement to zero. Shaded regions are pointwise 95 percent confidence bands constructed from 1,000 normal coefficient draws using the estimated robust covariance matrix, with the same normalization applied to each draw.

Figure A.10: Sensitivity to FPCA Implementation Choices

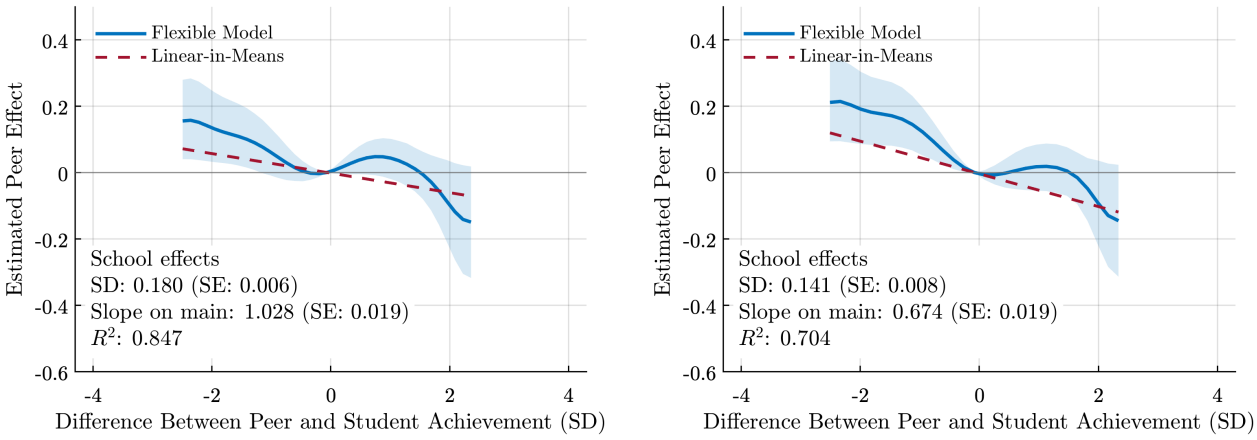


(a) Main specification (b) Four FPCs (c) Tail threshold: 1% (d) Tail threshold: 5%

(e) 50-point evaluation grid (f) 200-point evaluation grid (g) KDE bandwidth: 0.2 (h) KDE bandwidth: 0.4

Notes: Each panel reports pooled OLS estimates for sixth-grade mean MathELA achievement, with peer achievement measured relative to the focal student's baseline mean MathELA score. The blue curve is the FPCA peer-effect estimate, the blue shading is its pointwise 95 percent confidence interval, and the burgundy line is the corresponding linear-in-means estimate. Panel (a) uses three enrolled-density FPCs, ten expected-density FPC controls, a 2.5 percent threshold in each tail, a 100-point evaluation grid, and MATLAB's default KDE bandwidth. Panels (b)–(h) change only the setting named in the panel heading: four FPCs, a 1 or 5 percent tail threshold, a 50- or 200-point grid, or a KDE bandwidth of 0.2 or 0.4. Expected peer densities are averaged over 50 simulated assignments. All regressions include school and year fixed effects, student characteristics, and propensity scores; the FPCA specifications additionally include ten expected-density FPC coordinates, while the linear-in-means specifications include the expected peer mean. Student characteristics comprise cubic polynomials in baseline math and ELA scores and indicators for English-learner status, sex, race and ethnicity, poverty, and disability. In Panels (b)–(h), SD (SE) is the bias-corrected cross-school standard deviation under the alternative specification and its delta-method standard error. Slope (SE) and R^2 come from an unweighted regression of the alternative empirical-Bayes-shrunken school effects on the benchmark shrunken school effects, restricted to schools observed in both specifications.

Figure A.11: Peer Effects by Subject

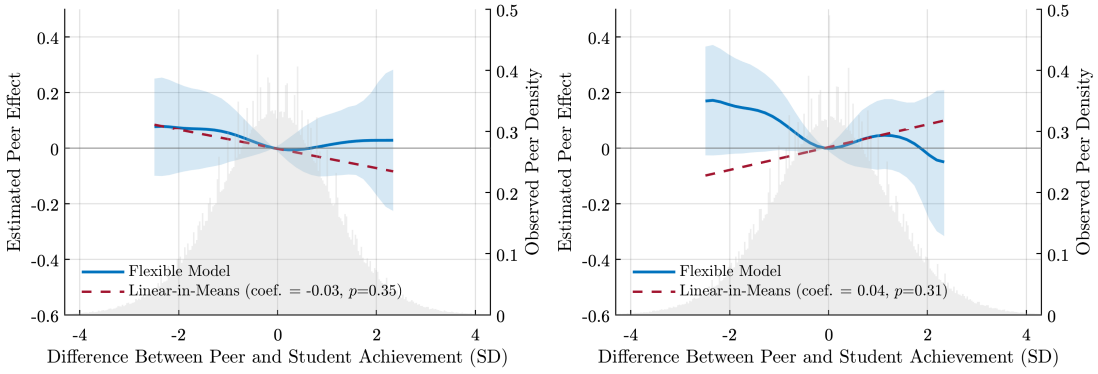


(a) Math scores

(b) ELA scores

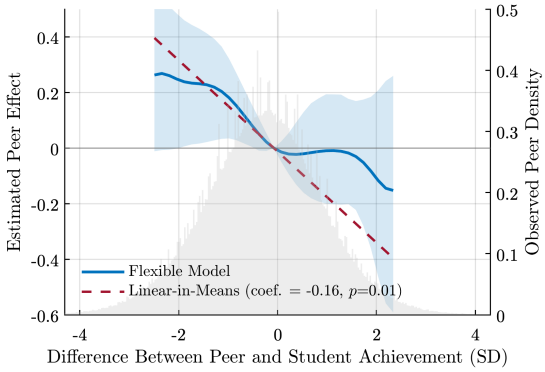
Notes: The panels use math and ELA, respectively, for both the sixth-grade outcome and baseline peer achievement. Peer achievement is measured relative to the focal student's baseline score in the same subject. All other choices are held at the main specification: three enrolled-density FPCs, ten expected-density FPC controls, a 2.5 percent threshold in each tail, a 100-point grid, MATLAB's default KDE bandwidth, school and year fixed effects, student characteristics, and propensity scores. The blue curve is the FPCA estimate, the blue shading is its pointwise 95 percent confidence interval, and the burgundy line is the linear-in-means estimate. The FPCA regressions control for the subject-specific expected-density coordinates, while the linear-in-means regressions control for the subject-specific expected peer mean. In each panel, SD (SE) is the bias-corrected cross-school standard deviation of the subject-specific school effects and its delta-method standard error. Slope (SE) and R^2 come from an unweighted regression of the subject-specific empirical-Bayes-shrunken school effects on the main specification's mean-MathELA shrunken school effects, restricted to schools observed in both specifications. The analysis reuses the previously computed math- and ELA-specific expected-density files and does not rerun the assignment simulations.

Figure A.12: Peer Effects by Student Sex and Race/Ethnicity

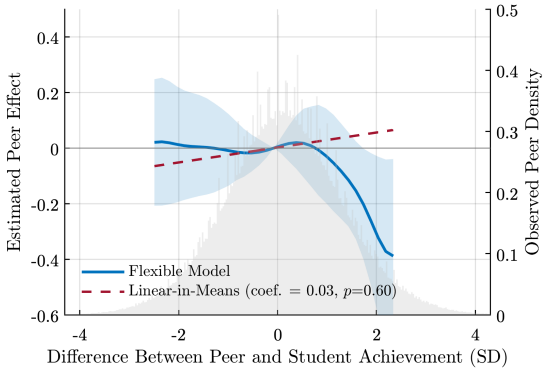


(a) Female students

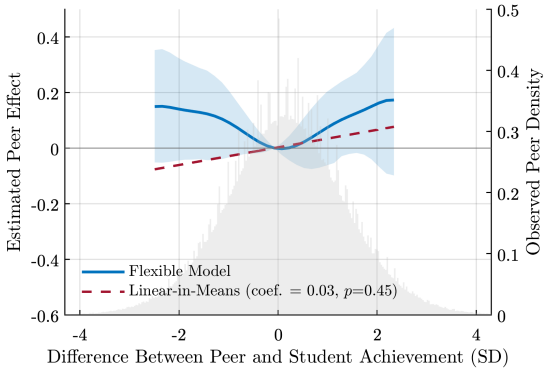
(b) Male students



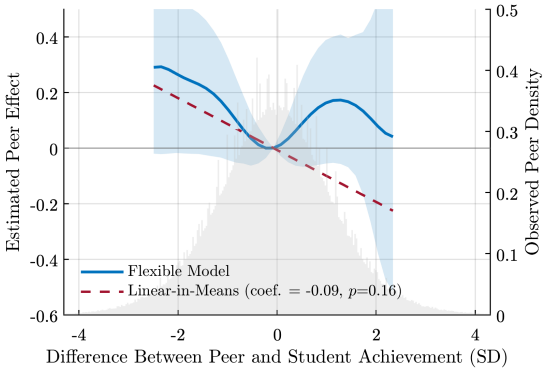
(c) Asian students



(d) Black students



(e) Hispanic students



(f) White students

Notes: Panels (a)–(f) report estimates for female, male, Asian, Black, Hispanic, and White students, respectively. Peer achievement is measured relative to the focal student's baseline achievement. The solid blue curve is the three-component FPCA estimate, the dashed burgundy line is the linear-in-means estimate, the blue band is a pointwise 95 percent confidence interval, and the gray histogram is the observed distribution of relative achievement among co-enrolled peers in each panel's sample. Both functions are normalized to zero at a relative peer score of zero. The models otherwise follow the main peer-effect specification.

B. Additional Tables

Table B.1: Predicted Achievement Effects of Shifting Relative Peer Distributions

	Flexible Model			Linear-in-Means		
	Full sample	High-achieving	Low-achieving	Full sample	High-achieving	Low-achieving
0.5 SD lower	0.031 (0.010)	0.052 (0.008)	0.015 (0.013)	0.024 (0.000)	0.053 (0.000)	0.006 (0.000)
0.5 SD higher	-0.029 (0.009)	-0.055 (0.006)	-0.011 (0.010)	-0.024 (0.000)	-0.053 (0.000)	-0.006 (0.000)

Notes: Entries report the average predicted change in achievement from shifting each student's observed relative peer-achievement distribution by 0.5 SD lower or higher. Predictions are computed separately using the flexible peer-effects model and the linear-in-means model, for the full sample and by baseline-achievement group. Parentheses report the standard deviation of predicted changes across students. Outcomes are measured in student-level standard deviations.

Table B.2: Moments of School Effects: Absolute Peer Achievement

	(1)	(2)	(3)
	Homogenous Effect Model	Heterogeneous Effect Model	
	Intercept	Intercept	High-Achieving Students
Mean	0	0	0.075
	-	-	(0.067)
SD	0.162	0.170	0.114
	(0.006)	(0.007)	(0.010)
Correlation: High-Achieving Students	-		-0.066
	-		(0.076)

Notes: The table reports moments of school effects on the standardized sixth-grade MathELA outcome in the pooled 2014–17 absolute-score analysis sample. Panel A estimates school effects jointly with three enrolled-peer FPCA coordinates and expected peer-distribution coordinates; Panel B omits peer-distribution controls. All models include school fixed effects, year fixed effects, school-assignment propensity scores, cubic polynomials in incoming math and ELA scores, and student demographic controls. In the heterogeneous model, the intercept is the school effect for low-achieving students (at or below the pooled median incoming MathELA score), and the high-achieving-student interaction component is the additional effect for high-achieving students. Baseline effects are normalized to have mean zero; the differential's mean is estimated. Standard deviations and correlations are corrected for fixed-effect estimation error. Parentheses report delta-method standard errors; dashes denote normalized or inapplicable cells.

Table B.3: Alternative Summaries of Peer Achievement

	FPCA	Peer variance		Mass at the Tail			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
FPCA coordinate 1	-0.162		-0.178		-0.254		-0.150
	(0.054)		(0.054)		(0.072)		(0.059)
FPCA coordinate 2	0.031		0.040		0.041		0.029
	(0.023)		(0.023)		(0.023)		(0.023)
FPCA coordinate 3	-0.146		-0.162		-0.202		-0.138
	(0.038)		(0.038)		(0.047)		(0.041)
Share peers in bottom 5%				-0.055	-0.235		
				(0.072)	(0.085)		
Share peers in top 5%				-0.023	0.119		
				(0.075)	(0.092)		
Share peers in bottom 1%						0.284	0.134
						(0.160)	(0.167)
Share peers in top 1%						-0.145	-0.003
						(0.210)	(0.214)
Variance of peer scores		-0.034	-0.048				
		(0.017)	(0.017)				
SD of school effects	0.162	0.154	0.163	0.154	0.161	0.155	0.162
SE of SD	0.006	0.005	0.006	0.005	0.006	0.005	0.006
N	143,147	143,147	143,147	143,147	143,147	143,147	143,147
R^2	0.808	0.807	0.808	0.807	0.808	0.807	0.808

Notes: The dependent variable is the average of standardized sixth-grade Math and Reading test-scores in the pooled 2014–17 relative-score analysis sample. Columns report models using three FPCA coordinates, bottom- and top-tail peer shares, peer-score variance, or combinations of these summaries. Each realized peer statistic enters jointly with its simulated expected counterpart; FPCA specifications additionally include expected peer-distribution coordinates. All columns include school fixed effects, year fixed effects, school-assignment propensity scores, cubic polynomials in incoming math and ELA scores, and demographic controls. Expected-peer controls are not displayed. Parentheses contain heteroskedasticity-robust coefficient standard errors.

Table B.4: Peer Effects by Own- and Peer-Achievement Buckets

	(1)	(2)	(3)	(4)
High-achieving peers x Low-achieving students	-0.105*			
	(0.041)			
High-achieving peers x High-achieving students	-0.129**			
	(0.042)			
Peer tercile 2 x Low-achieving students		0.069		
		(0.050)		
Peer tercile 3 x Low-achieving students		-0.094+		
		(0.053)		
Peer tercile 2 x High-achieving students		-0.164**		
		(0.055)		
Peer tercile 3 x High-achieving students		-0.158**		
		(0.053)		
High-achieving peers x Own tercile 1			-0.114*	
			(0.045)	
High-achieving peers x Own tercile 2			-0.113**	
			(0.043)	
High-achieving peers x Own tercile 3			-0.127**	
			(0.045)	
Peer tercile 2 x Own tercile 1				0.090
				(0.059)
Peer tercile 3 x Own tercile 1				-0.109+
				(0.058)
Peer tercile 2 x Own tercile 2				-0.054
				(0.055)
Peer tercile 3 x Own tercile 2				-0.107*
				(0.054)
Peer tercile 2 x Own tercile 3				-0.174**
				(0.065)
Peer tercile 3 x Own tercile 3				-0.163**
				(0.056)
SD of school effects	0.160	0.159	0.161	0.159
SE of SD	0.006	0.006	0.006	0.006
N	143147	143147	143147	143147
R^2	0.808	0.808	0.808	0.808

Notes: The dependent variable is the standardized sixth-grade MathELA outcome in the pooled 2014–17 absolute-score analysis sample. Columns (1)–(4) use 2-by-2, 2-by-3, 3-by-2, and 3-by-3 own-by-peer achievement buckets, respectively. Own achievement is divided at the median or into terciles. Peer-score halves are constructed by summing decile shares 1–5 and 6–10; peer terciles use the corresponding tercile shares. Models interact realized and expected peer shares with the student's own-achievement bucket. The lowest peer-score group—low-achieving peers or peer tercile 1—is omitted in every specification. All columns include own-bucket indicators, school and year fixed effects, school-assignment propensity scores, cubic polynomials in incoming math and ELA scores, and demographic controls. Expected-share interactions are not displayed. The school-effect SD is corrected for fixed-effect estimation error, and its SE is computed by the delta method. Parentheses contain heteroskedasticity-robust coefficient standard errors. +, ∗, and ∗∗ denote significance at the 10, 5, and 1 percent levels, respectively.

C. Interpretation of $\beta(r)$

Let

$$\mathcal{P}(f) \equiv \int \beta(r) f(r)\, dr$$

denote the peer component of predicted achievement under peer-achievement density f. For exposition, suppose that the support of f is partitioned into n bins of common width Δr, with bin centers $r_1, \ldots, r_n$. The peer component can then be approximated by the Riemann sum

$$\mathcal{P}(f) \approx \Delta r \sum_{l=1}^{n} \beta(r_l) f(r_l).$$

Consider two bins, indexed by j and k, with centers r_j and r_k. Let g denote a counterfactual peer-achievement density that increases the density in bin k by Δf and decreases the density in bin j by the same amount:

$$\begin{aligned} g(r_k) &= f(r_k) + \Delta f, \\ g(r_j) &= f(r_j) - \Delta f, \\ g(r_q) &= f(r_q) \qquad \text{for all } q \notin \{j, k\}. \end{aligned}$$

Provided that $\Delta f \leq f(r_j)$, this counterfactual transfers probability mass $m \equiv \Delta f \times \Delta r$ from the bin centered at r_j to the bin centered at r_k.

The difference between the Riemann-sum approximations to the peer component under g and f is:

$$\mathcal{P}(g) - \mathcal{P}(f) \;\approx\; \Delta r \left[\beta(r_k)\Delta f - \beta(r_j)\Delta f\right] \;=\; m \left[\beta(r_k) - \beta(r_j)\right].$$

D. FPCA representation and estimation details

This appendix provides additional details on the FPCA representation used to estimate the peer effects model. Most of the arguments in this appendix closely follow Chang et al. (2023). We proceed in three steps: (i) explain why uncorrelated loadings imply no exact multicollinearity in our specification, (ii) explain why truncating the FPCA expansion does not induce omitted-variable

bias from approximation error in the linear functional regression setting, and (iii) show why FPCA loadings are uncorrelated.

D.1 Why FPCA truncation does not create omitted-variable bias in the linear model

Consider the linear functional regression model

$$y_i = \langle \beta, f_i \rangle + \varepsilon_i,$$

where $E[\varepsilon_i \mid f_i] = 0$ and f_i is centered. Let $\{v_k\}_{k \geq 1}$ denote the FPCA basis (eigenfunctions of the covariance of f_i) and define the loadings (scores) $c_{ik} := \langle v_k, f_i \rangle$ and coefficients $\beta_k := \langle \beta, v_k \rangle$. By the FPCA expansion,

$$f_i = \sum_{k \geq 1} c_{ik} v_k, \qquad \beta = \sum_{k \geq 1} \beta_k v_k,$$

so that

$$\langle \beta, f_i \rangle = \sum_{k \geq 1} \beta_k c_{ik}.$$

Our empirical specification uses only the first m loadings as regressors, i.e., it replaces $\langle \beta, f_i \rangle$ by its truncated approximation $\sum_{k=1}^{m} \beta_k c_{ik}$. The corresponding approximation residual (the omitted term) is

$$u_i := \sum_{\ell > m} \beta_\ell c_{i\ell}.$$

In a standard finite-dimensional regression, truncation would generate omitted-variable bias if u_i were correlated with the included regressors $\{c_{ik}\}_{k \leq m}$. Under FPCA, however, distinct loadings are uncorrelated: for $k \neq \ell$, $E[c_{ik} c_{i\ell}] = 0$. Therefore, for each $k \leq m$,

$$E[c_{ik} u_i] = \sum_{\ell > m} \beta_\ell \, E[c_{ik} c_{i\ell}] = 0.$$

Thus the approximation residual is orthogonal to the retained components, so truncation does not induce omitted-variable bias in the linear functional regression setting. This argument relies on linearity and does not generally extend to nonlinear functional models; see Chang et al. (2023).

D.2 Why FPCA loadings are uncorrelated

We restrict attention to square-integrable functions and assume that both the functional parameter $\beta(\cdot)$ and the peer distributions $(f_i)_i$ belong to a Hilbert space H endowed with inner product

$$\langle \beta, f \rangle := \int \beta(r) f(r)\, dr.$$

FPCA constructs the basis $\{v_k\}_{k=1}^{\infty}$ as eigenfunctions of the covariance operator of $(f_i)_i$. For centered functions, the loading (score) for component k is $c_{ik} := \langle v_k, f_i \rangle$, and standard FPCA results imply

$$E[c_{ik} c_{i\ell}] = 0 \text{ for } k \neq \ell, \qquad E[c_{ik}^2] = \lambda_k.$$

Hence the loadings are mutually uncorrelated across retained components.

Using the first m components, we approximate

$$f_i \approx \sum_{k=1}^{m} \langle v_k, f_i \rangle v_k.$$

Define the mapping $\pi : H \to \mathbb{R}^m$ by

$$\pi(f) := (\langle v_1, f \rangle, \ldots, \langle v_m, f \rangle)'.$$

Under this representation, the peer-effects term satisfies

$$\langle \beta, f_i \rangle \approx \pi(\beta)' \pi(f_i),$$

which leads to the finite-dimensional approximation of the structural equation. Given an estimate $\widehat{\pi(\beta)}$, the functional parameter is recovered as

$$\hat{\beta}(\cdot) := \sum_{k=1}^{m} \widehat{\pi(\beta)}_k v_k(\cdot).$$

Under regularity conditions, $\hat{\beta}(\cdot)$ consistently estimates the projection of $\beta(\cdot)$ onto the span of $\{v_k\}_{k=1}^{m}$ (Chang et al., 2023).

E. Details of NYC School Choice Process & Data Cleaning

E.1 NYC School Admission

Depending on the eligibility criteria, middle schools are classified into three types—district schools, borough schools, and citywide schools. The city is divided into 5 boroughs and 32 community school districts. A student's residence or elementary school determines eligibility for each type of school. In the academic year 2014–2015, among 670 programs, 14 were citywide programs, 39 were borough programs, and the rest were district programs.

Middle school programs use a variety of admission methods—Unscreened, Limited Unscreened, Screened, Screened: Language, Zoned, and Talent Test. Unscreened programs admit students by a random lottery number, and Limited Unscreened programs use rules that give priority to those who attend information sessions or open houses. Screened programs as well as Screened: Language programs select students by individually assessed measures such as elementary school GPA, statewide test scores, punctuality, and interviews. Zoned programs guarantee admission or give priority to students who reside in the school's zone, and Talent Test programs use auditions as the main criteria.

The timeline of the admission process is as follows. By December, students are required to submit their school application list. By March, DA algorithms are run that determine students' assignments. Students who accept their offer finalize, and if a student rejects an offer she goes to the next round. This describes the main round of the entire system. A majority of students finalize in the main round (about 85% each year). Students who are not assigned in the main round or reject the assignment go to the Supplementary round, which is organized similarly to the main round and includes schools/programs that did not fill their capacities in the main round, or programs that are newly opened. Finally, there is an administrative round in which students who are not assigned a school even after the second round are administratively assigned to a school.

E.2 Sample Construction and Data Cleaning

Our starting population consists of 216,525 students who participated in the New York City middle school application process in the 2014–15, 2015–16, and 2016–17 cohorts. We first drop students without baseline characteristics or 5th grade baseline test scores, which leaves 201,602

observations. We then drop students with missing assigned program identifiers or missing enrolled program identifiers, leaving 151,620 observations. Next, we drop five students who are assigned to a school without peers, leaving 151,615 observations. Finally, because our outcome is 6th grade achievement, we drop students without 6th grade test scores, leaving us with 146,972 observations. The placebo balance tests use a smaller sample, 137,771 observations, because they additionally require non-missing 4th grade test scores. To use a consistent sample across analyses, we use 137,771 observations as our final estimation sample.

F. Asymptotic rates and the N-versus-J trade-off

This appendix derives the asymptotic rates under which school and peer effects can be consistently estimated in the two-step procedure described in Section 3. We consider a sequence of markets indexed by n; the number of schools J_n and the arrays of school effects introduced below may change across markets,[1] and each school $j = 1, \ldots, J_n$ enrolls $q_n := N_n/J_n$ students.

As the number of schools J_n grows in a market with N_n students, a trade-off emerges. Consistent estimation of school-level effects would require enough students per school, $N_n/J_n \to \infty$. However, when schools become large, realized peer composition converges to its expected value, reducing the variation in the school-level peer-composition regressors. Thus, the two requirements pull J_n in opposite directions.

Under a stylized design, we show that the two consistency conditions are compatible. When school and peer effects are common across student types, both peer-adjusted school and peer effects can be consistently estimated when the number of schools satisfies

$$\sqrt{N_n} \ll J_n \ll N_n,$$

i.e., $J_n/\sqrt{N_n} \to \infty$ and $N_n/J_n \to \infty$. When effects are type-specific, the same rates apply under the mildly strengthened condition that the number of schools, J_n, grows more slowly than $N_n/\log N_n$, which guarantees that every school contains students of every type.

Before stating the formal conditions and propositions, let us provide intuition on the condition above. $N_n/J_n \to \infty$ is required for the average student outcomes at school level, a main object in the

[1] That is, we work in a triangular-array asymptotic framework.

first step of our identification argument, to be precise. The second-step regressor, peer composition, has variance proportional to $J_n/N_n = 1/q_n$, because each school-level peer distribution averages over q_n students. While variance therefore shrinks as each school becomes larger, the number of second-step observations, J_n, also grows with n. The total second-step regressor variation is therefore of order $J_n(J_n/N_n) = J_n^2/N_n$, so the variance of the peer-effect coefficient is of order N_n/J_n^2 and vanishes if $J_n^2/N_n \to \infty$.

Below, we formalize this claim as Proposition 1 and Remark 2 shows that its rate restrictions also become necessary under generic conditions. Throughout, we work in a benchmark where assignment coincides with enrollment.

Environment: We allow all effects to differ by the student's own type. Let $t_i \in \mathcal{T}$ denote student i's baseline type, where $\mathcal{T}$ is a fixed finite set. A type may combine the student's own baseline-score category and demographic group (= $(\tau_i, g(i))$ in Section 3). The type contains the own characteristics used to construct the peer-composition measures below.

Let $\widetilde{h}(t) \in \mathbb{R}^M$ collect the contribution of a student of type t to peer composition, such as evaluations of M basis functions or score-bin indicators after omitting one redundant category. Define its population demeaned version

$$h(t) := \widetilde{h}(t) - E[\widetilde{h}(t_i)],$$

where the expectation is taken over the population distribution of student types. Let $j(i)$ denote the school that student i attends; school j's peer-composition vector is

$$c_{jn} := \frac{1}{q_n} \sum_{i:\, j(i)=j} h(t_i).$$

Student outcomes satisfy

$$y_i = \alpha^{t_i} + \mu_{jn}^{t_i} + \left(b^{t_i}\right)' c_{jn} + \varepsilon_i, \qquad j = j(i), \tag{6}$$

where ε_i is the student's unobserved ability. For each type t, define $\delta_{jn}^t := \mu_{jn}^t + (b^t)' c_{jn}$.

Assumption 3 (Balanced random-admission design)**.**

(i) $J_n \to \infty$, *every school has size* $q_n = N_n/J_n$, *and* M *and* $|\mathcal{T}|$ *are fixed. Each type has a positive population share:* $\pi_{\min} := \min_{t\in\mathcal{T}} \Pr(t_i = t) > 0$.

(ii) Random admission places students into schools independently of student characteristics, so that the pairs (t_i, ε_i) *are i.i.d. across all students, within and across schools.*

(iii) For every type t, *the school effects are fixed before admission and satisfy*

$$\frac{1}{J_n}\sum_{j=1}^{J_n} \mu_{jn}^t = 0, \qquad \frac{1}{J_n}\sum_{j=1}^{J_n} (\mu_{jn}^t)^2 \le C.$$

In (iii), the zero-mean restriction is a normalization: a common shift in $\{\mu_{jn}^t\}_{j=1}^{J_n}$ can be absorbed by α^t. The bound on the second moment, by contrast, is substantive, preventing the cross-school dispersion of school effects from diverging.

Assumption 4 (Regularity conditions)**.** *For every type* t,

$$E[\varepsilon_i \mid t_i = t] = 0, \qquad E[\varepsilon_i^2 \mid t_i = t] \le C, \qquad E[\varepsilon_i^4 \mid t_i = t] \le C.$$

The centered vector h *satisfies* $E[h(t_i)] = 0$ *by construction; we assume that*

$$\Sigma_h := E[h(t_i)h(t_i)']$$

is positive definite. Because $\mathcal{T}$ *is finite,* h *is bounded, so all higher moments of* h *are finite.*

The first condition on ε_i is only an orthogonalization: any nonzero effect of own type on the outcome level is absorbed into the type intercept α^{t_i}.[2]

Two-step Identification Argument: Let $q_{jtn} := \#\{i : j(i) = j,\ t_i = t\}$ be the number of type-t students in school j. The first step computes school-by-type cell means:

$$\bar{y}_{jn}^t := \frac{1}{q_{jtn}} \sum_{i:\, j(i)=j,\ t_i=t} y_i. \tag{7}$$

[2]In the homogeneous case introduced below, where α^t is constant across types, the same condition is instead substantive: a student's own type must not shift her outcome beyond the common intercept.

In the second step, $\bar{y}^t_{jn}$ is regressed on an intercept and c_{jn} separately for each type. We also consider the homogeneous special case in which the outcome parameters do not vary with the student's own type ($\alpha^t = \alpha$, $\mu^t_{jn} = \mu_{jn}$, and $b^t = b$), in which case the first step instead uses the school mean including all types and the second step is run once.

We now fix a type t and suppress its superscript. Let $\bar{c}_n := J_n^{-1} \sum_j c_{jn}$, $\bar{y}_{\cdot n} := J_n^{-1} \sum_j \bar{y}_{jn}$, and $\widetilde{c}_{jn} := c_{jn} - \bar{c}_n$. The second-step estimators are

$$\begin{aligned} \widehat{b}_n &:= \left(\sum_{j=1}^{J_n} \widetilde{c}_{jn} \widetilde{c}'_{jn} \right)^{-1} \sum_{j=1}^{J_n} \widetilde{c}_{jn} (\bar{y}_{jn} - \bar{y}_{\cdot n}), \\ \widehat{\alpha}_n &:= \bar{y}_{\cdot n} - \widehat{b}'_n \bar{c}_n, \qquad \widehat{\mu}_{jn} := \bar{y}_{jn} - \widehat{\alpha}_n - \widehat{b}'_n c_{jn}, \\ \widehat{\delta}_{jn} &:= \widehat{\mu}_{jn} + \widehat{b}'_n c_{jn} = \bar{y}_{jn} - \widehat{\alpha}_n. \end{aligned} \tag{8}$$

Proposition 1 (Rate conditions for school and peer effects). *Let Assumptions 3 and 4 hold.*

(a) (Homogeneous effects) Suppose α^t, μ^t_{jn}, and b^t do not vary with t, and the first step uses full-school means. Then for every fixed school j,

$$\begin{aligned} \widehat{\delta}_{jn} - \delta_{jn} &= O_p(q_n^{-1/2}) + O_p(J_n^{-1}), \\ \widehat{b}_n - b &= O_p\left(\sqrt{\frac{q_n}{J_n}} \right) + O_p(J_n^{-1/2}), \\ \widehat{\mu}_{jn} - \mu_{jn} &= O_p(q_n^{-1/2}) + O_p(J_n^{-1/2}). \end{aligned}$$

Aggregate and peer-adjusted school effects are consistently estimated when $q_n \to \infty$, and peer-effect coefficients when $q_n / J_n \to 0$. Both requirements hold when

$$1 \ll q_n \ll J_n, \qquad \textit{equivalently,} \qquad \sqrt{N_n} \ll J_n \ll N_n.$$

(b) (Type-specific effects) In model (6), suppose in addition that $q_n / \log J_n \to \infty$. Then the same three rates hold for $\widehat{\delta}^t_{jn}$, $\widehat{b}^t_n$, and $\widehat{\mu}^t_{jn}$, for every fixed type t and school j. All requirements hold when

$$\log J_n \ll q_n \ll J_n, \qquad \textit{equivalently,} \qquad \sqrt{N_n} \ll J_n \ll N_n / \log N_n.$$

The proof is given in Section F.1.

Remark 1 (Type-specific versus homogeneous effects)**.** *The condition, $J_n \ll N_n/\log N_n$, ensures that all these cell sizes are sufficiently large simultaneously, so that no first-step mean is based on an unusually small number of observations (Lemma 1). This is a uniform cell-size condition rather than an additional statistical-rate condition: conditional on all cells being sufficiently large, the O_p rates are unchanged. Because a type-t cell contains approximately $\pi_t q_n$ students, its first-step mean has variance of order $\frac{1}{\pi_t q_n}$, which differs from the homogeneous case only by the fixed factor $\frac{1}{\pi_t}$. In polynomial regimes with $J_n = N_n^a$ and $1/2 < a < 1$, we have $q_n = N_n^{1-a}$, and therefore $q_n/\log J_n \to \infty$ automatically. Thus, the admissible polynomial region remains $1/2 < a < 1$, equivalently $\sqrt{N_n} \ll J_n \ll N_n$. The homogeneous model, whose first step pools all q_n students within a school, does not need a similar condition (part (a)).*

Remark 2 (Sharpness)**.** *The preceding rate restrictions are also necessary under standard nondegeneracy conditions. If the conditional variance of ε_i is bounded below, bounded q_n prevents the first-step mean error from vanishing. If $J_n^{-1}\sum_j(\mu_{jn}^t)^2$ is bounded away from zero, then the part of $\widehat{b}_n^t - b^t$ that results from school effects has variance of order q_n/J_n, so it does not generally vanish if $q_n/J_n \not\to 0$. Finally, in the type-specific case, if $q_n = O(\log J_n)$ with a sufficiently small constant, some school-by-type cell is empty with probability approaching one, so the cell means are not even well defined in every school.*

F.1 Proof of Proposition 1

We first show that each school-by-type cell is well populated, a condition on which we will rely in the type-specific case. For a constant $0 < \underline{\pi} < \pi_{\min}$, define an event

$$A_n := \left\{ \min_{j \le J_n,\, t \in \mathcal{T}} q_{jtn} \;\ge\; \underline{\pi}\, q_n \right\},$$

on which every school-by-type cell mean is based on a sample of order q_n.

Lemma 1 (Cell coverage)**.** *Under Assumption 3, if $q_n/\log J_n \to \infty$, then $\Pr(A_n) \to 1$ for any $0 < \underline{\pi} < \pi_{\min}$.*

Proof. By Assumption 3(ii), the types of the q_n students enrolled in any given school are i.i.d. draws from the population type distribution, so q_{jtn} is a sum of q_n i.i.d. indicators with success

probability $\Pr(t_i = t) \geq \pi_{\min}$. Hoeffding's inequality with $\epsilon := \pi_{\min} - \underline{\pi} > 0$ gives

$$\Pr\big(q_{jtn} < \underline{\pi} q_n\big) \leq \Pr\big(q_{jtn} < (\Pr(t_i = t) - \epsilon)\, q_n\big) \leq e^{-2\epsilon^2 q_n}.$$

The event A_n^c occurs if at least one of the $J_n|\mathcal{T}|$ school-by-type cells is too small, so a union bound yields

$$\Pr(A_n^c) \leq \sum_{j \leq J_n,\, t \in \mathcal{T}} \Pr\big(q_{jtn} < \underline{\pi} q_n\big) \leq J_n|\mathcal{T}|\, e^{-2\epsilon^2 q_n} \to 0,$$

since $q_n / \log J_n \to \infty$. □

Proof of Proposition 1. We prove the statements for a fixed type t in the type-specific model and suppress the superscript t throughout. All statements are established on the event A_n: in the type-specific case, Lemma 1 gives $\Pr(A_n) \to 1$ under $q_n / \log J_n \to \infty$, so restricting attention to A_n does not affect O_p statements. In the homogeneous case, all students in a school form a single pooled cell of deterministic size q_n, so no coverage condition is needed and A_n may be ignored; the argument below then applies verbatim and delivers part (a). The proof proceeds in three steps: Step 1 bounds the noise in the first-step means, $\bar{y}_{jn}^t$, Step 2 bounds error in the peer coefficient, $\hat{b}_n$, and Step 3 bounds error in the (peer-effect-adjusted) school effect estimates, $\hat{\mu}_{jn}$.

Step 1. Substituting the model (6) into the cell mean (7) gives

$$\bar{y}_{jn} = \alpha + \delta_{jn} + u_{jn}, \qquad u_{jn} := \frac{1}{q_{jtn}} \sum_{i:\, j(i)=j,\; t_i = t} \varepsilon_i,$$

so the first-step error u_{jn} is the average of the residuals of the q_{jtn} type-t students in school j. Conditional on the realized types $\{t_{i'} : j(i') = j\}$, the composition vector c_{jn} is fixed, and by Assumption 3(ii) and 4, each residual has conditional mean zero given all realized types in the school:

$$E[\varepsilon_i \mid \{t_{i'} : j(i') = j\}] = E[\varepsilon_i \mid t_i] = 0,$$

hence $E[u_{jn} \mid \{t_{i'} : j(i') = j\}] = 0$. The residuals are independent across students with variance bounded by C (Assumptions 3(ii) and 4), so the variance of their average is at most C / q_{jtn}. On A_n,

conditional on $q_{jtn}, q_{jtn} \geq \underline{\pi} q_n$, hence

$$E[u_{jn} \mid q_{jtn}] = 0, \qquad E[u_{jn}^2 \mid q_{jtn}] \leq \frac{C}{\underline{\pi}\, q_n}.$$

For any $K > 0$, Chebyshev's inequality gives

$$\Pr\big(|u_{jn}| > K q_n^{-1/2}\big) \leq \frac{C}{\underline{\pi} K^2},$$

which is arbitrarily small for K large. In other words,

$$\bar{y}_{jn} - (\alpha + \delta_{jn}) = u_{jn} = O_p(q_n^{-1/2}).$$

A standard moment calculation for averages of independent, mean-zero terms similarly yields the fourth-moment bound $E[u_{jn}^4 \mid q_{jtn}] \leq C/q_n^2$, which we use in Step 2.

Step 2. Let C_n denote the $J_n \times M$ matrix with jth row $\widetilde{c}_{jn}'$, and stack the school effects and first-step errors as $\mu_n := (\mu_{1n}, \ldots, \mu_{J_n n})'$ and $u_n := (u_{1n}, \ldots, u_{J_n n})'$. Substituting $\bar{y}_{jn} = \alpha + \mu_{jn} + b' c_{jn} + u_{jn}$ into the definition of $\widehat{b}_n$ in equation (8) gives the exact decomposition

$$\widehat{b}_n - b = (C_n' C_n)^{-1} C_n' \mu_n + (C_n' C_n)^{-1} C_n' u_n.$$

Two facts recur below. First recall that $c_{jn} = q_n^{-1} \sum_{i:j(i)=j} h(t_i)$, and by Assumption 4, this is an average of q_n mean-zero and bounded contributions. Thus,

$$E[c_{jn}] = 0, \qquad E[c_{jn} c_{jn}'] = \frac{\Sigma_h}{q_n}, \qquad c_{jn} = O_p(q_n^{-1/2}).$$

Second, by Step 1, u_{jn} is likewise mean zero with variance of order q_n^{-1}: both the regressor and the noise shrink at the same $q_n^{-1/2}$ scale. Moreover, quantities attached to different schools involve disjoint sets of students, whose characteristics are i.i.d. by Assumption 3(ii), and are therefore independent across schools.

Denominator. Each $c_{jn} c_{jn}'$ has expectation Σ_h / q_n, so $C_n' C_n = \sum_j \widetilde{c}_{jn} \widetilde{c}_{jn}'$ concentrates around

$J_n \Sigma_h / q_n$ by a law of large numbers across schools. Hence

$$(C_n' C_n)^{-1} = O_p\left(\frac{q_n}{J_n}\right).$$

School-effect term. By the normalization $\sum_j \mu_{jn} = 0$ in Assumption 3(iii), the recentering drops out exactly: $C_n' \mu_n = \sum_j \widetilde{c}_{jn} \mu_{jn} = \sum_j c_{jn} \mu_{jn}$. The school effects are fixed before admission while the compositions are mean zero and independent across schools, so each summand has mean zero and the summands are independent. The variance of the sum is therefore at most

$$\sum_j \mu_{jn}^2 \left\| E[c_{jn} c_{jn}'] \right\| \leq C J_n \cdot \frac{C}{q_n},$$

using $J_n^{-1} \sum_j \mu_{jn}^2 \leq C$ from Assumption 3(iii). Chebyshev's inequality then gives

$$C_n' \mu_n = O_p\left(\sqrt{\frac{J_n}{q_n}}\right).$$

Noise term. Each summand of $\sum_j c_{jn} u_{jn}$ is the product of two mean-zero within-school averages, each of order $q_n^{-1/2}$. Because c_{jn} is fixed conditional on the realized types while $E[u_{jn} \mid \{t_{i'} : j(i') = j\}] = 0$ (Step 1), we have $E[c_{jn} u_{jn}] = E\big[c_{jn} \, E[u_{jn} \mid \{t_{i'} : j(i') = j\}]\big] = 0$ despite the overlap in students. By the Cauchy–Schwarz inequality together with the fourth-moment bounds on c_{jn} and u_{jn}, each product has second moment of order q_n^{-2}. Summing J_n mean-zero, independent products,

$$\sum_j c_{jn} u_{jn} = O_p\left(\frac{\sqrt{J_n}}{q_n}\right);$$

the recentering correction $\sum_j (\widetilde{c}_{jn} - c_{jn}) u_{jn} = -J_n \bar{c}_n \bar{u}_n = O_p(q_n^{-1})$ is of lower order, so $C_n' u_n = O_p(\sqrt{J_n}/q_n)$ as well.

Combining. Multiplying the two numerator orders by $(C_n' C_n)^{-1} = O_p(q_n / J_n)$,

$$\widehat{b}_n - b = O_p\left(\sqrt{\frac{q_n}{J_n}}\right) + O_p\left(J_n^{-1/2}\right).$$

The first term reflects the fixed school effects acting as the error term of the second-step regression;

the second reflects first-step estimation noise.

Step 3. Let $\bar{u}_n := J_n^{-1} \sum_j u_{jn}$; averaging J_n mean-zero cell errors of variance $O(q_n^{-1})$ gives $\bar{u}_n = O_p((J_n q_n)^{-1/2})$, and likewise $\bar{c}_n = O_p((J_n q_n)^{-1/2})$. The second-step intercept then obeys

$$\widehat{\alpha}_n - \alpha = \bar{u}_n - (\widehat{b}_n - b)'\bar{c}_n = O_p((J_n q_n)^{-1/2}) + O_p(J_n^{-1}),$$

where the second rate combines Step 2 with the order of $\bar{c}_n$. For the school effects,

$$\widehat{\mu}_{jn} - \mu_{jn} = (u_{jn} - \bar{u}_n) - (\widehat{b}_n - b)'(c_{jn} - \bar{c}_n).$$

Step 1 bounds the first difference by $O_p(q_n^{-1/2})$. Since $c_{jn} - \bar{c}_n = O_p(q_n^{-1/2})$, Step 2 bounds the second difference by $O_p(J_n^{-1/2})$. Hence

$$\widehat{\mu}_{jn} - \mu_{jn} = O_p(q_n^{-1/2}) + O_p(J_n^{-1/2}).$$

Finally,

$$\widehat{\delta}_{jn} - \delta_{jn} = \left(\bar{y}_{jn} - \alpha - \delta_{jn}\right) - (\widehat{\alpha}_n - \alpha) = u_{jn} - (\widehat{\alpha}_n - \alpha) = O_p(q_n^{-1/2}) + O_p(J_n^{-1}).$$

□

G. Testing the Identifying Assumption

This appendix extends the specification test for the identifying assumption from Angrist et al. (2017, 2024) to our model with both school and peer effects. We first state the infeasible test when we do not approximate the peer distributions. We then describe the finite-dimensional test relying on FPCs.

As discussed in Section 3, lottery-induced school assignment variation generates instrumental variables for both school and peer assignments. But we rely on the aRC-CIA to estimate the model to avoid problems from many weak IVs. Instead, we use IVs to conduct over-identification tests à la Sargan (1958).

Infeasible Test: Recall that the lottery generates two corresponding assignment shocks. The school assignment shock is

$$z_{ij}^{S} = d_{ij} - p_{ij}, \tag{9}$$

where d_{ij} is the realized school assignment and p_{ij} is the expected school assignment. The realized school assignment is a binary indicator for whether student i is assigned to school j, and the expected school assignment is the probability of being assigned to school j given the lottery design.

Similarly, the assigned-peer shock is the function

$$z_i^P(r) = f_i^{\text{asg}}(r) - \widetilde{f}_i^{\text{asg}}(r), \tag{10}$$

where $f_i^{\text{asg}}(r)$ is the realized assigned-peer distribution and $\widetilde{f}_i^{\text{asg}}(r)$ is the expected assigned-peer distribution.

Under aRC-CIA, they are orthogonal to the structural residual:

$$\mathbb{E}\left[\varepsilon_i \begin{pmatrix} z_i^S \\ \\ z_i^P \end{pmatrix}\right] = 0. \tag{11}$$

Equation (11) provides the basis for the specification test. Intuitively, once the model has accounted for students' observed characteristics and ex ante assignment risk, lottery-induced assignment surprises should not systematically predict residuals from the structural model in Equation (1).

Feasible Test: Testing Equation (11) is infeasible since peer distributions are functions, not finite-dimensional objects, and we do not observe ε_i. We therefore use the same functional principal components $v_1(r), \ldots, v_M(r)$ as in the main analysis to summarize these objects, and replace the unobserved residuals with the fitted values from Equation (4).

Note that the residual ϵ_i in Equation (4) is the sum of the structural residual and approximation error:

$$\epsilon_i^M = \varepsilon_i + u_i^M,$$

where u_i^M collects the approximation error from retaining only M components.

To further make the assigned-peer shock in Equation (10) finite dimensional, we likewise project it onto the retained components:

$$z_{im}^P = \int \left[f_i^{\text{asg}}(r) - \widetilde{f}_i^{\text{asg}}(r) \right] v_m(r)\, dr, \qquad m = 1, \ldots, M. \tag{12}$$

Now, let Z^S collect the school-assignment shocks and let Z^P collect the M peer-assignment shocks in Equation (12). In the empirical implementation, we aggregate school shocks into groups to avoid a test with many weak moments. Specifically, following Angrist et al. (2024), we classify schools into ventiles using the shrunken school effect estimates from running Equation (4) and then sum the school shocks within each ventile. This gives a total of 20 school-assignment shocks. Let

$$Z = \begin{pmatrix} Z^S & Z^P \end{pmatrix}$$

denote the resulting finite-dimensional shock matrix.

The feasible moment condition is

$$\mathbb{E}[Z_i \epsilon_i^M] = \mathbb{E}[Z_i \varepsilon_i] + \mathbb{E}[Z_i u_i^M] = 0. \tag{13}$$

aRC-CIA, along with model specification, implies the first equality $\mathbb{E}[Z_i \varepsilon_i] = 0$. The test also requires the retained shocks to be orthogonal to the truncation error, $\mathbb{E}[Z_i u_i^M] = 0$. The finite-dimensional procedure should therefore be interpreted as a joint test of aRC-CIA, the structural model specification, and the adequacy of the FPCA approximation.

Test Statistic and Decomposition: Let W collect the nuisance controls included in Equation (4), and let $P_{Z|W}$ project onto the portion of the shock space that is orthogonal to W. This residualization ensures that the test uses lottery-induced variation not already explained by the controls. We estimate Equation (4) on the full analysis sample and evaluate the validation moments on the testing subsample with meaningful school-assignment and peer-composition risk. Specifically, we restrict the testing subsample to students who face a nondegenerate assignment probability at at least one school on their submitted list and whose total variation distance between the realized

and expected assigned-peer densities is at least 0.014, the median applicant-level TVD reported in Table 2.

To test Equation (13), we run the following equation:

$$\epsilon^M = Z_i\gamma_1 + W_i\gamma_2 + \xi_i, \tag{14}$$

and test $\gamma_1 = 0$. The omnibus test statistic is

$$\widehat{T} = \frac{\left(Y - \widehat{Y}^M\right)' P_{Z|W} \left(Y - \widehat{Y}^M\right)'}{\widehat{\sigma}^2_\xi}, \qquad \widehat{\sigma}^2_\xi = \frac{1}{N_T}\sum_{i\in\mathcal{T}}(\widehat{\xi}_i)^2, \tag{15}$$

where $\widehat{Y}^M$ denotes the fitted values obtained using only the peer and school effect part of Equation (4) and $\mathcal{T}$ denotes the testing subsample and N_T is its size. Under the null, homoskedasticity, and standard regularity conditions, $\widehat{T}$ has an asymptotic chi-squared distribution with degrees of freedom equal to the rank of the residualized shock matrix.

The decomposition closely follows Angrist et al. (2024). Following the forecast-bias approach, define

$$\widehat{\phi} = \left(\widehat{Y}^{M\prime} P_{Z|W} \widehat{Y}^M\right)^{-1} \widehat{Y}^{M\prime} P_{Z|W} Y. \tag{16}$$

This coefficient measures the relationship between realized outcomes and model predictions along the assignment-shock variation. If the model is correctly specified, its predictions should be correctly scaled, so $\widehat{\phi}$ should be close to one. Adding and subtracting $\widehat{\phi}\widehat{Y}^M$ in the numerator of Equation (15) gives the exact decomposition

$$\widehat{T} = \underbrace{\frac{(\widehat{\phi} - 1)^2 \widehat{Y}^{M\prime} P_{Z|W} \widehat{Y}^M}{\widehat{\sigma}^2_\xi}}_{\text{forecast bias}} + \underbrace{\frac{\widehat{e}'_\phi P_{Z|W} \widehat{e}_\phi}{\widehat{\sigma}^2_\xi}}_{\text{overidentification}}, \tag{17}$$

where

$$\widehat{e}_\phi = Y - \widehat{\phi}\widehat{Y}^M.$$

The first component tests whether the forecast coefficient equals one. The second asks whether the assignment shocks explain any residual variation after allowing the fitted values to be rescaled by

$\widehat{\phi}$. The forecast-bias component has one degree of freedom, and the overidentification component has one fewer degree of freedom than the omnibus statistic.

School and peer components: The overidentification component can be further allocated between the school- and peer-assignment shocks. If the column spaces of Z^S and Z^P were orthogonal after residualizing the nuisance controls, their contributions would add uniquely. In general, however, the two shock blocks can overlap, so their allocation therefore depends on which block enters the projection first.

For example, placing the school shocks first gives

$$P_{Z|W} = P_{Z^S|W} + P_{Z^P|(W,Z^S)}. \tag{18}$$

The first projection assigns the variation shared by the two blocks to the school shocks, while the second assigns to the peer shocks only their incremental variation. Reversing the order gives

$$P_{Z|W} = P_{Z^P|W} + P_{Z^S|(W,Z^P)}. \tag{19}$$

We report a Shapley-style allocation that averages the two orderings. In particular, the school contribution is

$$\widehat{T}_{\text{school}} = \frac{1}{2\widehat{\sigma}_\xi^2}\left[\widehat{e}_\phi' P_{Z^S|W}\widehat{e}_\phi + \widehat{e}_\phi' P_{Z^S|(W,Z^P)}\widehat{e}_\phi\right], \tag{20}$$

and the peer contribution is

$$\widehat{T}_{\text{peer}} = \frac{1}{2\widehat{\sigma}_\xi^2}\left[\widehat{e}_\phi' P_{Z^P|W}\widehat{e}_\phi + \widehat{e}_\phi' P_{Z^P|(W,Z^S)}\widehat{e}_\phi\right]. \tag{21}$$

These two quantities sum to the overidentification component in Equation (17). They are diagnostic allocations, rather than separate test statistics. The decomposition results are shown in Table 3.

Interpretation: A rejection of the omnibus test does not isolate a single source of misspecification. It may indicate that the aRC-CIA assumption fails, that the finite-dimensional FPCA approximation is inadequate, or that the outcome model omits relevant heterogeneity in school or peer effects. We therefore examine the stability of the test across alternative specifications. These include varying

the number of retained FPCA components and allowing school and peer effects to differ across student groups. Stability across these specifications makes it less likely that a rejection is driven solely by a particular approximation or heterogeneity restriction, although it cannot establish the identifying assumption by itself.

Table G.1: Identification Test

	(1) LIM	(2) $M = 4$	(3) Heterogeneity
Forecast coefficient	1.052	1.045	1.039
	(0.045)	(0.045)	(0.043)
First-stage F-statistic	469.96	411.49	185.33
Bias tests			
Omnibus test (χ^2 d.f.)	26.70 (21)	33.41 (24)	47.65 (46)
	(0.181)	(0.096)	(0.405)
Forecast bias	1.33	1.01	0.80
	(0.249)	(0.314)	(0.371)
Overidentification (χ^2 d.f.)	25.37 (20)	32.40 (23)	46.85 (45)
	(0.188)	(0.092)	(0.396)
Overidentification decomposition			
School effect	20.50	26.31	40.35
Peer effect	4.87	6.09	6.50
N (testing)	43,007	43,007	43,007
N (full model)	143,147	143,147	143,147

Notes: This table repeats the validation tests in Table 3 under alternative specifications. Column (1) replaces the FPCA peer measure with mean relative peer achievement and controls for its expected assigned counterpart. Column (2) retains four rather than three enrolled-peer FPCA components. Column (3) allows school and peer effects to differ for students above and at or below the pooled median of baseline achievement. All specifications control for school-assignment propensity scores, cohort effects, student characteristics, and the appropriate expected-peer measure; the testing sample is the same as in Table 3. Parentheses beneath the forecast coefficient contain its standard error; parentheses beneath test statistics contain p-values, while degrees of freedom appear beside the omnibus and overidentification statistics. The school and peer rows report the Shapley decomposition of the overidentification statistic.

H. Numerical Optimization Details

Grid and integration. We evaluate all densities on the 55-point achievement grid used to construct the FPCA basis and integrate with the rectangle rule, the rule used to normalize the estimated densities. On this grid we remove from each basis function its small constant component, so that each v_m^a sums to exactly zero. Every candidate density then has the same total mass for any choice of loadings, matching the continuum property that the basis functions integrate to zero.

Solution and global optimality. The reformulation over FPCA loadings does not by itself guarantee a convex problem, because the curvature of the peer term depends on the estimated peer-effect functions. Under our estimates, the quadratic form that defines the peer-effect component is negative definite, so the objective is concave in the loadings; with linear constraints, the continuous problem is a convex program and any local optimum is global. We solve it with MATLAB's interior-point algorithm.

Channel extremes. The peer component is concave, so its maximum is also global. Minimizing the peer component is not a convex problem: we report the lowest value found across several starting points (0.020 SD below the observed assignment), and a Lagrangian bound implies that no feasible allocation lowers the peer component by more than 0.024 SD. The school component is linear in the loadings, so its extremes are exact, but many allocations attain them. Among these, we report the one with the largest peer component, computed by maximizing the peer component subject to the school component lying within 0.0001 SD of its extreme.

Validity restrictions. In the implementation, we also require each implied low-achieving share to lie between zero and one, and each implied density to be no lower on the grid than the lowest value among the FPCA reconstructions of the observed school densities (-0.067). With three components, due to FPCA approximation error, some observed reconstructions dip slightly below zero, so exact nonnegativity would exclude observed compositions. Both restrictions are affine in the loadings and hold at every observed school, so they hold throughout the convex hull and do not bind.

Cohorts. The main exercise pools the three cohorts (502 schools, 151,293 students), which we interpret as a stationary policy applied to each cohort. Solving the problem separately for each

cohort yields similar gains.