The Value of a Peer*

Ingo E. Isphording

(Max Planck Institute for Research on Collective Goods)

Albert Thieme

(University of Zurich)

Ulf Zölitz

(University of Zurich)

July, 2026

Abstract

This paper introduces peer value-added, a new approach to quantify the contribution of individual peers to the performance of others. Using data with repeated random assignments of students to university sections, we estimate the distribution of peer effects. Having a peer with a one standard deviation higher value-added increases performance by 0.013 standard deviations. Peer value-added predicts out-of-sample student performance in cross-validations. Observable peer characteristics, including prior achievement, are poor predictors of peer value-added. These findings show that canonical peer effects models, which measure peer influence exclusively based on observable characteristics, fail to capture the full extent of social spillovers.

Keywords: peer effects, peer value-added, spillovers

JEL classification: I21, I24, J24

* We thank our valuable peers Steffen Altmann, David Dorn, Jan Feld, John N. Friedman, Daniel Hamermesh, Jonas Radbruch, Nicolás Salamanca, Mikko Silliman, Michela Tincani and seminar participants at the NBER education meeting, the 10th IWAEE Workshop in Catanzaro, the CESifo Area Conference on Economics of Education, the EALE conference, DICE, the University of Chicago, the University of Zurich, IZA, the Arne Ryde Workshop on Economics of Education in Lund, the Bristol Workshop on Economic Policy Interventions and Behaviour, the University of Padova, the Università Cattolica del Sacro Cuore, the University of Frankfurt, and the Ski and Labor Seminar in Engelberg for providing helpful comments that contributed to this paper. Maximilian Mähr, Anna Valyogos, Matthew Bonci, Timo Haller and Jeffrey Yusof provided outstanding research assistance. We gratefully acknowledge financial support from the University of Zurich URPP Equality of Opportunity. Ingo E. Isphording: Max Planck Institute for Research on Collective Goods. Albert Thieme: University of Zurich. Ulf Zölitz: University of Zurich, Department of Economics and Jacobs Center for Productive Youth and Child Development, Schönberggasse 1, 8001 Zurich. ulf.zoelitz@econ.uzh.ch.

1. Introduction

The canonical definition of a peer effect is any externality arising from the presence of one peer instead of another (Epple and Romano, 2011; Sacerdote, 2011). However, this definition does not align with empirical practice in the social sciences. The dominant approach is to project outcomes on peers' observable characteristics, which captures only a subset of potential spillovers and leaves the overall variation in peer quality unknown.

In this paper, we introduce a new approach that quantifies the importance of peers more directly: Peer value-added (PVA) captures an individual's overall contribution to the performance of others. It summarizes *any* spillover on others' performances, regardless of whether spillovers stem from observed or unobserved peer characteristics. Peer value-added can be estimated even when information on peer characteristics is absent. Our key parameter of interest is the standard deviation of the distribution of peer value added. This parameter captures how much peers matter in a given environment.

We estimate peer value-added in a European business school in which students are repeatedly and randomly assigned to sections of, on average, 12 peers. This repeated randomization plays a dual role: On the one hand, it ensures exogeneity: the presence of any particular peer in a section is independent of the characteristics of the other students in that group. On the other hand, repeated assignment of the same peer across many different sections provides the variation necessary to identify that peer's unique contribution to their classmates' performance. In this environment, we estimate peer value-added by adopting the three-step empirical Bayes framework by Walters (2024). First, we regress student performance on a high-dimensional set of peer indicators. This delivers noisy but unbiased raw estimates of the value of each peer and a corresponding standard error. Second, we use these raw peer value-added estimates and their standard errors to estimate the unobserved distribution of latent peer value-added (also known as mixing distribution). The estimate of the standard deviation of this distribution is our key parameter to summarize how much peers matter. Third, we obtain empirical Bayes PVA measures by shrinking the raw peer value-added estimates toward the estimated mean of the unobserved peer value-added distribution.

We document five main results. First, peer value-added—the ability to improve fellow students' performance—differs significantly between peers. Being randomly assigned to a peer with a one standard deviation higher value-added increases a student's grade by 0.013 SD. Replacing a bottom decile peer with a top decile peer improves performance by 0.046 SD.

Second, peer value-added is a valid predictor of student performance in new out-of-sample social interactions. In a cross-validation exercise, we show that leave-out (jackknife) empirical Bayes PVA estimates are forecast-unbiased predictors of student performance in interactions that were not part of the initial estimation sample. Empirical Bayes PVA estimates thus contain meaningful information about students' expected spillovers. This validation mitigates concerns about overfitting and is a prerequisite for deriving valid policy recommendations from peer effects estimates.[1]

Third, the quality of a student's *entire* peer group has a moderate effect on academic performance. "Lucky" students—assigned to sections in the top quartile of section-average peer quality—score, on average, 0.057 SD higher than "unlucky" students in bottom-quartile sections. Peer value-added not only varies substantially at the individual level, but being assigned to a section of more valuable peers has a meaningful impact on student performance.

Fourth, observable peer characteristics, including peer achievement, are poor predictors of peer value-added. This result suggests that the multitude of studies that estimate the effect of having higher-achieving peers miss substantial spillovers.

Fifth, high-value-added peers seem to influence performance by improving the classroom atmosphere and increasing student effort. We find that exposure to more valuable peers improves group functioning, the quality of peer-to-peer interactions, and increases students' self-study hours. These results indicate that students respond to the presence of valuable peers in the classroom.

With this paper, we contribute both conceptually and empirically to the peer effects literature. Conceptually, we focus on the person-specific peer value-added, which differs from the current state-of-the-art approach in the peer effects literature that

[1] Establishing the out-of-sample predictive power is essential as Carrell, Sacerdote and West (2013) show that standard peer effects estimates can be invalid performance predictors after reshuffling students.

relates a group-level aggregate of a single observable characteristic to student outcomes. This literature studies how different types of peers, for example, higher-achieving peers, black peers, female peers, or free-lunch peers, affect student performance.[2] In contrast, we propose a more comprehensive approach to estimating peer effects that recognizes that spillovers can arise from both observed and unobserved peer characteristics. This approach closely follows canonical definitions of peer effects—any externality arising from the presence of one peer instead of another (Epple and Romano, 2011, Sacerdote, 2011)—without relying on a prior about which observables can produce peer effects. It complements recent papers that quantify individual contributions to *team production* in laboratory experiments (Weidmann and Deming 2021), sports (Arcidiacono et al. 2017), and scientific production (Bonhomme 2021). In contrast to these studies, we identify peer contributions to *individually determined* performance.

Empirically, we contribute to a more nuanced understanding of peer effects in education by providing the first estimates of the peer value-added distribution and person-specific peer value-added estimates. The weak relationship between peer value-added and observable peer characteristics indicates that social spillovers are not always well captured by student observables.

Our approach to identifying peer value-added is inspired by the literature on teacher value-added.[3] This literature also finds that observable characteristics are poor predictors of teacher value-added. A comparison of the magnitude of value-added estimates shows that the variation in teacher value-added estimates is three to five times larger than the variation in peer value-added estimates. In other words, being assigned to a teacher who is predicted to be one standard deviation more effective has a larger effect on students' performance than being assigned to a peer who is predicted to be one standard deviation more valuable.

[2] Epple and Romano (2011) and Sacerdote (2011) provide comprehensive literature overviews on the peer effects literature. Booij, Leuven and Oosterbeek (2017), Carrell, Sacerdote and West (2013), and Garlick (2018) provide recent evidence on ability peer effects.

[3] Koedel et al. (2015) and Hanushek and Rivkin (2006) provide a comprehensive review on the extensive literature on teacher value-added.

Our peer value-added approach has several limitations. First, we assume linear-additive, peer-specific constant effects. In reality, peer effects may vary across contexts and individuals. Consequently, our peer value-added measures only identify the average, reduced-form social spillover of individual peers. They do not incorporate match-specific complementarities between peers and students that have been documented in the peer effects literature.[4] Second, our approach has demanding data requirements: It depends on repeated and ideally exogenous assignments of individuals to peer groups. This limits the number of settings in which peer value-added can be credibly estimated. Third, due to our finite sample size and limited number of reassignments, our *individual* empirical Bayes PVA measures and their standard deviation have limited precision. We therefore focus on the distribution of PVA and abstain from deriving strong policy implications from empirical Bayes PVA estimates of the value of individual peers.

2. Institutional Environment, Data, and Randomization

We estimate PVA with data from a European business school covering six academic years between 2009–10 and 2014–15. In this school, students are repeatedly and randomly assigned to study sections. In the following, we provide an overview of the institutional details, introduce the dataset, and provide evidence on the randomness of the section assignment process.[5]

2.1 Institutional Environment

The business school offers bachelor's, master's, and Ph.D. programs. Our analysis focuses on the two major bachelor’s programs in which all first-year students follow the same general course structure and take the same eight compulsory courses. In addition, a typical student takes 10–12 compulsory and elective courses in the second and third years and spends one term abroad.

[4] We test whether peer influence changes when students share characteristics such as gender, nationality, or similar past achievement. We find large overlap in value-added distributions for same versus other type peers and limited evidence supporting match-specific peer spillovers. However, in our setting, we lack the statistical power to provide conclusive evidence on such match effects.

[5] A similar description of the institutional details is provided in Elsner, Isphording and Zölitz (2021) and in Feld and Zölitz (2017).

In each course, students have weekly lectures and meet in sections of, on average, 12 students. For each course, students are assigned to a different section. These sections are the peer groups we analyze. Section meetings last for two hours and usually take place twice per week. Students work on their study material alone or in groups before section meetings where they discuss the material with their peers. These discussions are guided by an instructor, who can be a professor, lecturer, graduate student, or undergraduate student (Feld et al. 2018).

A key feature of the business school is that the scheduling office randomly assigns students and instructors within courses to sections. The random assignment is a by-product of the scheduling software used at the institution. We exclude from our analysis courses for which course coordinators did not comply with the standard section allocation procedure. For the remaining courses, neither students nor instructors interfere with the random allocation process. Since 2011, the random assignment has been stratified by nationality.[6] In Section 2.3 we present balancing checks that confirm the random assignment of students to sections.

The assignment of students to sections is binding. Students may switch from the assigned section only for medical reasons, or when the student is a top athlete and must attend sports practice during the schedule time. Students must attend their designated section, and instructors keep a record of attendance.[7] If students miss more than three meetings of their designated section, they are not admitted to the exam. First-year compulsory courses use one assessment only: a written exam at the end of each period, graded at the course level. Some later courses also base the final grade on additional components such as presentations, participation, and term papers.

2.2 Data

Data structure: Each observation is a student–course record. The data include the student's course-specific grade and demographics (age, gender, nationality). We also

[6] In about 5 percent of the student-section assignments, students are manually reassigned due to scheduling conflicts. To test whether this reassignment affects results, we include parallel course fixed effects for courses taken at the same time in the estimation as one of our robustness checks. These fixed effects have no meaningful impact on our results.

[7] The attendance data are not centrally stored and thus are not available to us.

observe student and section identifiers, which allow us to link each student to their peers in the same section.

Student performance: Our main outcome is students' final course grade. The business school uses a grading scale ranging from 1 to 10, with 5.5 being the lowest passing grade. To simplify the interpretation of our estimates, we standardize grades to have a mean of zero and a standard deviation of one over our estimation sample. End-of-course grades serve two purposes. First, they act as our primary outcome. Second, they allow us to compute pre-assignment student GPA based on all grades received before a section assignment takes place. This predetermined student GPA is, by construction, missing for the first teaching period.

Sample restrictions: We apply two sets of sample restrictions to arrive at our estimation sample. First, we define minimum requirements about which student-section-level observations we include: We exclude observations with missing grade information, as it is unclear whether the student attended any section meetings, and we exclude students who appear in less than three different sections. This restriction primarily excludes exchange students and students from other study programs, who typically enroll in fewer courses, and ensures sufficient variation in student exposure across sections.[8] Second, we define the set of students for whom we estimate peer value-added. Specifically, we include only students who enrolled between 2009 and 2012 in one of the two main bachelor's programs, which allows us to observe them through graduation, and who appear in at least eight different sections, corresponding to the number of compulsory first-year courses. Last, we drop sections in which more than 50 percent of students are ineligible for peer value-added estimation. This step avoids contamination of individual estimates from peers for whom we lack sufficient identifying variation. We apply this procedure iteratively, as dropping sections may cause previously eligible students to fall below the eight-section threshold. Together,

[8] This restriction also avoids overfitting in our later jackknife cross-validation exercise. If students appear in only one section, student fixed effects perfectly explain their outcomes.

these restrictions ensure that we retain a sample in which peers are observed across enough distinct group compositions to support credible estimation of peer value-added.

Descriptive statistics: Table 1 shows descriptive statistics for our estimation sample. We observe 3,970 students across seven cohorts: 38 percent are female, and students are, on average, 20 years old. Regarding nationality, 24 percent of students are Dutch and 47 percent are German. The average course grade is 6.57 on the 1–10 grading scale, with a standard deviation of 1.82. The pre-assignment GPA has a similar mean of 6.86. For our analysis, we standardize end-of-term course grades to have a mean of zero and a standard deviation of one across the estimation sample.

Panels C, D, and E provide summary statistics for courses, sections, and students.[9] We observe students in 360 courses, each containing, on average, 12.4 sections and 148.8 students. We observe 4,447 sections with an average of 12 students per section.

We estimate individual peer-value added for 2,692 of the 3,970 students in our sample after implementing the sample restrictions discussed above.[10] Through the repeated assignment, a student interacts, on average, with 138 students who meet the requirements for individual peer-value added estimation. Conditional on being assigned to the same section once, two students are assigned to the same section in, on average, 1.3 courses. In our estimation sample we observe one peer in on average 17.8 unique sections.

Due to the repeated random assignment of all students our data is characterized by a very thick network. Appendix Table A1 shows that 99 percent of the sample is connected via two links or less. Full connectedness is achieved via a maximum of three links. Figure A1 shows the distribution of the number of unique peers that each student

---

[9] Throughout the paper, we refer to a course as one course taught in a specific year. For example, *Introductory Microeconomics* in 2010 represents one course.

[10] There is a tradeoff between estimating peer-value added for as many students as possible and the quality of individual peer value-added estimates. Lowering the minimum number of sections leads to noisier individual peer-value–added measures and reduces their predictive power in out-of-sample validation. Increasing the minimum number of sections beyond eight does not substantially improve predictive power.

meets. The figure highlights that students meet many distinct peers—on average 103—throughout their studies.

2.3 Randomization Check

The random assignment of students to sections alleviates concerns about the potential nonrandom selection of students into peer groups. To confirm the randomness of section assignment, we regress student pre-treatment characteristics (previous GPA, age,[11] gender, and the rank of the student ID—a proxy for a student's tenure at the university) on section dummies. Wang (2010) first proposed this test, and it has more recently been used by Antonovics et al. (2022) to detect ability tracking within schools. Under random assignment, we would expect section dummies to *not* jointly predict pre-treatment characteristics. To check whether this is the case, we run one regression per course of student observables on section dummies. We then collect the p-values of F-tests for joint significance from these regressions. Under the null hypothesis of random assignment, these p-values should be uniformly distributed, and they should reject the null hypothesis as frequently as specified by the specific confidence level. For example, 5 percent of these p-values should be significant at the 5 percent level.

Figure A2 in the appendix shows histograms of p-values for the joint F-tests of group dummies for six student characteristics. The figure shows that the p-values are uniformly distributed for the dependent variables female, age, GPA, and ID rank, confirming that section dummies are unrelated to these characteristics. We also see that a large share of p-values for the dependent variables Dutch and German are close to one. For these student characteristics, we observe *less* variation than expected under random assignment. The result for Dutch and German nationality is mechanical due to the stratified assignment by nationality.

Table 2 provides summary statistics for these balancing tests and shows that the actual rejection rates are close to the rejection rates expected under random assignment. Specifically, p-values are significant at the 5 percent level in approximately 4 percent of the cases and significant at the 1 percent level in approximately 1 percent of the

[11] Age is missing for about 20 percent of the sample for which we impute it using the sample mean.

cases. Overall, the balance tests confirms that section assignment was random conditional on nationality.

3. Conceptual Framework and Estimation

Our goal is to estimate the distribution of peer value-added—the causal effect of an individual peer on a classmate's performance. We formalize PVA in a potential-outcomes framework and develop an empirical strategy that exploits repeated random reassignment of students to sections to identify its distribution. We then discuss two challenges—limited-mobility bias and high-dimensional parameter estimation.

3.1 Potential Outcome Framework

Consider student i's educational achievement when she is assigned to section s of course c. In section s, student i interacts with a peer group G_{ics}, defined as the set of classmates in i's section excluding herself.

$Y_{ics}(G_{ics})$ denotes the potential achievement outcome for student i in course c, assigned to section s, given the peer group G_{ics}. We begin by expressing the potential outcome as the sum of four noninteracting components:

$$Y_{ics}(G_{ics}) = \alpha_i + \delta_c + \Theta_{ics} + \epsilon_{ics}, \tag{1}$$

where α_i represents student i's individual ability, δ_c captures the effect of course c, and Θ_{ics} denotes the sum of peer spillovers that student i experiences because she is assigned to peer group G_{ics}. The error term ϵ_{ics} is assumed to be idiosyncratic. The central innovation of this paper is to disaggregate Θ_{ics} into individual peer-specific spillovers and estimate them directly. This requires the following two key assumptions:

Assumption (1) Peer-Specific Constant Effects: Each peer j has a unique peer spillover effect γ_j that is the same across all students $i \neq j$.

Assumption (2) Linear Additivity: The total peer effect on student i (Θ_{ics}) is the sum of the individual effects from each peer j in peer group G_{ics}.

Peer-specific constant effects imply that each peer has the same impact on any student she meets and in every context in which she is observed. With this assumption, we follow teacher-value–added models, in which the aim is to capture the portable, person-specific contribution rather than contingent, match-specific dynamics.

By assuming linear additivity we treat the section-level peer effect as the sum of the distinct influences of each student in the section. This assumption mirrors classic educational production function approaches (Todd and Wolpin 2003; Ben-Porath 1967; Hanushek, 1979) and is an implicit assumption of canonical linear-in-means models. Linear additivity is also a key assumption of two-way fixed effects approaches that explain wage inequality in labor economics (Abowd et al. 1999).

Together, these assumptions provide the simplest workable framework for isolating and quantifying peer spillovers in settings in which each student is repeatedly assigned to different peers. These assumptions would be violated if peer effects are match-specific. In Section 4.3 we test whether peer influence depends on shared characteristics and find little evidence for match-specific effects. Nevertheless, peer effects may vary across contexts and individuals. In this case, peer value-added only captures the average, reduced-form social spillover of individual peers. They do not incorporate match-specific complementarities between peers and students that the peer effects literature has documented. Under random assignment we expect potential deviations from constant effects to be orthogonal to the identities of individual peers.[12]

Under Assumptions (1) and (2) we can decompose the total peer spillover effect into the sum of individual spillovers:

$$\Theta_{ics} = \sum_{j \in G_{ics}} \gamma_j. \tag{2}$$

Combining equations (1) and (2), the full potential outcome equation then becomes

[12] The linear additivity assumption implies that one student affects every classmate in the same way, independent of the presence of other peers. This is implausible in very large classes. In our setting, however, classes are generally small—12 students on average—and students interact intensively with each other.

$$Y_{ics}(G_{ics}) = \alpha_i + \delta_c + \sum_{j \in G_{ics}} \gamma_j + \epsilon_{ics}. \qquad (3)$$

Under the additivity and the constant-effects assumption, we define γ_j as the individual causal effect of peer j:

$$\gamma_j = Y_{ics}(G_{ics}) - Y_{ics}\left(G_{ics}^{-j}\right), \qquad (4)$$

where G_{ics}^{-j} represent the peer group G_{ics} excluding peer j. The parameter γ_j thus captures the total contribution of peer j on students in section s, aligning closely with the canonical definitions of peer effects that incorporate any effect resulting from the presence of one peer instead of another (Epple and Romano, 2011) or any externality arising from the presence of a peer (Sacerdote, 2011).

3.2 Intuition of Estimating Peer Value-Added

To estimate the distribution of peer value-added, we face two challenges. First, we must isolate peer effects from other sources of variation, such as students' own ability, course-specific impacts, and classroom-level shocks. Second, we need to disentangle the effect of each peer from those of the other in the same group.

We address both challenges by leveraging that in our setting, students are *repeatedly and randomly assigned* to small peer groups. This repeated randomization plays a dual role. First, the randomization ensures exogeneity: the presence of any particular peer in a student's group is independent of the student's own characteristics and of the characteristics of other peers in that group. Second, the repeated assignment of the same student across many different group compositions provides the variation necessary to identify that student's unique contribution to their peers' performance.

The core intuition is simple: Suppose we want to estimate the peer value-added of a student named John. In one course, John is randomly placed in a section that performs well. This alone does not tell us much; perhaps the section benefited from a particularly good instructor or some other favorable classroom shock. However, if John is repeatedly assigned to different sections and his presence consistently correlates with better outcomes for his classmates, we can infer that John himself contributes positively to his peers' performance. Since the assignment process is random and other influences are averaged out across many rounds of reassignment, the consistent performance boost can be attributed to John's individual peer value-added.

While the logic is straightforward, in finite samples each individual peer effect is estimated noisily, so the cross-sectional variation of raw peer value-added estimates overstates true heterogeneity. What we care about is not the individual estimates per se, but the amount of genuine variation in peer influence. Our target parameter is the between-peer standard deviation of peer value-added rather than the student's individual effects. We therefore use an empirical Bayes approach that uses the noisy raw estimates and their standard errors to estimate the underlying distribution of true peer value-added. Additionally, we recover estimates of each individual peer's predicted effect on student performance. These empirical Bayes PVA estimates shrink the unbiased but noisy raw estimates toward the mean of the estimated PVA distribution—with more shrinkage for noisier estimates—yielding more-reliable predictions of a peer's effect in new group assignments.

3.3 Estimation Strategy

In the following, we describe the empirical strategy to estimating peer value-added in an empirical Bayes framework. After isolating variation plausibly attributable to peer interactions, we follow the three-step empirical Bayes approach by Walters (2024). First, we estimate raw peer value-added estimates and their standard errors. Second, we use these raw estimates to estimate the latent distribution of peer value-added (fitting a normal prior), thereby correcting for sampling noise. This provides us with an estimate of our key parameter of interest: the standard deviation of the latent distribution of peer value-added. Finally, we construct empirical Bayes estimates of PVA by shrinking the noisy raw estimates toward the mean of the latent distribution.

Isolating variation. We start by isolating the variation in student achievement that can conceivably be attributed to peer interactions.[13] To this end, we regress the student performance Y_{ics} on student fixed effects μ_i and course fixed effects λ_c.

$$Y_{ics} = \mu_i + \lambda_c + u_{ics}. \tag{5}$$

From this regression, we obtain the residuals $\hat{u}_{ics}$, which capture the variation in Y_{ics} that is not explained by student ability and course effects. These residuals capture the part of student performance that can be affected by peers.

Estimating raw peer value-added. Next, we regress the residual student achievement $\hat{u}_{ics}$ on a large set of dummy variables indicating the presence of each individual peer j in G_{ics} (i.e., which peers share section s with student i).

$$\hat{u}_{ics} = \beta_0 + \sum_{j \neq i} \gamma_j D_{ijcs} + \varepsilon_{ics}. \tag{6}$$

The coefficients γ_j capture the impact of the presence of peer j in section s on student performance in course c.[14] We estimate equation (6) with OLS and cluster our standard errors at the student and course level.[15] This regression gives us the raw peer value-added estimates $\hat{\gamma}_j$ and their standard errors.

[13] This approach is similar to the first step in the two-step approach of estimating teacher value-added (Chetty et al. 2014a, b). The adoption of a two-step approach is motivated by its computational efficiency and its capacity to leverage both within- and between-student variation (Bacher-Hicks and Koedel, 2023). As Walters (2024) observes, the mean residual or two-step approach resembles an uncorrelated random effects model. This model assumes that the mean characteristics of a group are uncorrelated with the value-added of an entity—in this case, the peer group. In our study setting, this assumption is appropriate due to the random assignment in groups.

[14] Note that equation (6) describes a regression of individual achievement on the *presence* of a peer, not their behavior. Current or past performance of a peer is not part of our estimation, which is important to avoid the reflection problem described by Manski (1993).

[15] Our many binary treatments of interest—peers—are assigned at the course level. However, clustering only at the course level alone would ignore serial correlation of grades within students (driven by, for example, motivation or unobserved ability) and risks understating standard errors when student-specific shocks matter. Clustering at the student level alone, however, would ignore common shocks within a course (for example, grading polices or an unusually tough final exam). Because these two cluster structures are cross-classified and non-nested, conventional one-way clustering will miss part of the

Because students are randomly assigned to sections, these raw estimates are unbiased estimates of individual-level PVA. Formally, student i's section is not systematically related to unobserved factors in the error term, such as potential common shocks at the section level $\mathrm{E}[\varepsilon_{ics} | D_{ijcs}] = 0$, where $D_{ijcs} = 1$ if peer j is in student i's study section s.

Estimating the variance of peer value-added (PVA): We assume individual PVA to follow a hierarchical model: The true individual PVA γ_j are drawn from a common but unknown normal distribution, commonly referred to as the mixing distribution (Walters, 2024):

$$\gamma_j \sim \mathcal{N}(\mu_\gamma, \sigma_\gamma^2), \tag{7}$$

where μ_γ is the overall mean peer effect, and our target parameter σ_γ^2 is the variance of the true PVA.

Our input for estimating this variance are the raw peer value-added estimates, which we assume to be unbiased estimates of the true γ_j, but also reflect normally distributed sampling error with a variance equal to the square of its standard error.

$$\hat{\gamma}_j \mid \gamma_j, s_j \sim \mathcal{N}(\gamma_j, s_j^2). \tag{8}$$

A consequence of this noise is that the empirical distribution of the raw estimates has a larger variance than the variance of the true effect. Our goal is therefore to deconvolve (filter out) this noise from the empirically observed distribution of $\hat{\gamma}$ to estimate the hyperparameters $(\mu_\gamma, \sigma_\gamma^2)$ of the distribution of PVA. In other words, we decompose

---

dependence structure and generally understate standard errors (Cameron, Gelbach and Miller 2011). Two-way clustering addresses both sources simultaneously and, by construction, reduces to the correct one-way variance if one source of correlation is in fact absent. Appendix A2.1 provides a series of robustness checks on how clustering choices affect key parameters of interest.

the total variance of the raw PVA measures into the variance of the true effect and variance due to sampling error. We estimate the mean of the underlying distribution as

$$\hat{\mu}_{\gamma} = \frac{1}{J}\sum_{j=1}^{J} \hat{\gamma}_j, \tag{9}$$

and its variance as

$$\hat{\sigma}_{\gamma}^2 = \frac{1}{J}\sum_{j=1}^{J} \left[\left(\hat{\gamma}_j - \hat{\mu}_{\gamma}\right)^2 - s_j^2\right], \tag{10}$$

where s_j^2 is estimated as the standard error of the estimate $\hat{\gamma}_j$. By subtracting s_j^2, we reduce the excess variance in the estimates due to noise, isolating the variance of the true peer effects σ_{γ}^2 as the core target parameter of our study.

Constructing empirical Bayes PVA estimates: In the last step, we obtain more-reliable estimates of the effect of an individual peer by combining the noisy raw estimates $\hat{\gamma}_j$ with the estimated moments of the empirical mixing distribution via empirical Bayes (EB) shrinkage. We multiply raw estimates with a measure of their reliability:

$$\hat{\gamma}_j^* = \left(\frac{\hat{\sigma}_{\gamma}^2}{\hat{\sigma}_{\gamma}^2 + s_j^2}\right)\hat{\gamma}_j + \left(\frac{s_j^2}{\hat{\sigma}_{\gamma}^2 + s_j^2}\right)\hat{\mu}_{\gamma}. \tag{11}$$

This method shrinks an individual's raw peer value-added estimate toward the overall mean value across peers, with weights depending on the reliability of the estimate, as determined by s_j^2. The weight $\frac{\hat{\sigma}_{\gamma}^2}{\hat{\sigma}_{\gamma}^2 + s_j^2}$ of an individual's observed value-added $\hat{\gamma}_j$ becomes smaller the larger its individual estimated sampling variance s_j^2 is, while the weight $\frac{s_j^2}{\hat{\sigma}_{\gamma}^2 + s_j^2}$ of the sample average increases with a larger sampling variance s_j^2. Thus, observed values of value-added will be shrunk more toward the sample average the larger their sampling variance is, reducing the influence of potentially noisy individual estimates.[16] The resulting empirical Bayes PVA estimates provide more reliable predictors of the causal effect of an individual peer's value-added for new social interactions.

[16] Once we estimate peer value-added, we subtract the overall peer-level mean (0.00031 which is almost zero). This implies that PVA is always interpreted relative to an average peer with a PVA of zero.

3.4 Estimation Challenges

Limited mobility bias: In the worker and firm effects literature (Abowd et al. 1999; Card et al. 2013), limited mobility bias arises when workers do not sufficiently move across firms. This limited movement can bias firm estimates, as the variation in worker mobility is insufficient to disentangle firm-specific components from other factors. We might face a similar limited mobility problem if students lack exposure to different peers or only move within disconnected peer sets.[17]

An important feature in our setting that differs from classical AKM analyses is the constant reassignment of all students to fresh sections in each period. Drawing an analogy to the AKM framework, our setting resembles one in which all workers are effectively randomized into new firms in every period for which we observe outcomes. This mitigates the issue of limited mobility bias. To show this more formally, Appendix Table A1 provides evidence on students' connectedness: 55.3 percent of students are connected via one common peer, 99.2 percent of students are connected via two nodes, and all students are connected via a maximum of three nodes. Taken together, the high connectedness of students suggests that limited mobility bias is not a major concern in our setting.

High-dimensional parameter estimation: While we can rely on a strong interconnectedness in our sample, we estimate a large number of parameters and have substantial overlap between peers. This introduces several interrelated problems. First, with many parameters and limited reassignments per peer, individual effects and standard errors will be noisy, affecting the quality of the deconvolution procedure. Second, the risk of overfitting may be amplified not only due to the number of parameters but also due to potential overlap in peer assignments, as multiple peer

[17] There are situations in which PVA cannot be computed or becomes unreliable. These situations typically involve insufficient variation or mobility in student-peer assignments. Examples include scenarios with no student mobility (students never change peer groups), single-student peer groups (each peer matched with exactly one student), disconnected networks (separate peer groups without cross-links), very few student reassignments (resulting in highly uncertain estimates), or a short observation window with minimal peer transitions. In these cases, peer effects become inseparable from individual student fixed effects, causing PVA estimates to be unidentifiable or overly noisy.

indicators explain each outcome. As a result, the model may capture idiosyncrasies in the data rather than true underlying peer effects.

We show below that concerns about overfitting appear limited: PVA is forecast-unbiased and predicts out-of-sample spillovers in new social interactions that were not part of the estimation sample. We also provide results on how well we can recover the variance of peer value-added from individual effects and standard errors in a simulated setting mirroring our actual settings in terms of section size, number of courses, and initial cohort size (Appendix A2.2). The results show that two counteracting parameters affect the precision of our estimates: A higher number of reassignments increase precision, while a larger peer group reduces precision. These simulations suggest that our data is well-suited to properly estimate our main target parameter.

4. Results

In the following section, we present our results. In Section 4.1, we establish that peers matter and describe the variance in their value-added. In Section 4.2, we validate our empirical Bayes PVA estimates by demonstrating predictive power in out-of-sample social interactions. In Section 4.3, we study whether observable characteristics predict PVA. In Section 4.4, we examine underlying mechanisms: self-reported learning from peers, group functioning, and study efforts.

4.1 Existence of Peer Effects and Variation in Peer Value-Added

Establishing that peers matter: We begin by testing for the existence of peer effects in the most general way: by estimating a linear model of student achievement that includes a full set of peer dummies—indicators for the presence of each peer in a section—and testing their joint significance. Table 3 reports the results. Across all specifications, the *F*-tests decisively reject the null hypothesis that peer identities have no effect on student outcomes.

Panel A shows results for the full set of peer indicators. The joint test strongly rejects the null ($p < 0.001$), regardless of whether standard errors are clustered at the student or section level. In Panel B, we cluster standard errors simultaneously at the student and course level. Because the number of peer indicators exceeds the number of clusters, we cannot directly conduct an *F*-test in the overall sample. We therefore partition peers into eight equally sized blocks, stratified by student ID (to avoid contamination by cohort effects), and test joint significance block by block. For each block, the hypothesis that peers have no effect is strongly rejected.

Panel C embeds the joint F-test in a permutation exercise. Here, we repeatedly draw random subsets of 359 peers, test their joint significance, and report average results across 1,000 repetitions. With the average p-value always below 0.001, Panel C confirms that peers systematically influence student achievement.

Taken together, Table 3 decisively rejects the hypothesis that peer identities are irrelevant across all specifications. Assigned peers thus clearly affect student performance.

The distribution of PVA: We next examine how much individual peers differ in their impact on performance. Figure 1 shows three distributions. The blue histogram shows the distribution of raw PVA estimates from equation (6). These estimates reflect the average performance change associated with exposure to a given peer. Because these raw estimates also reflect sampling error, they exaggerate the true variation in peer effects. The dashed black line plots the prior distribution of latent PVA, inferred from equations (9) and (10), which adjusts the raw distribution for sampling error. Finally, the red histogram shows the distribution of empirical Bayes PVA estimates obtained by applying empirical Bayes shrinkage to the raw estimates, as described in equation (11).[18]

Which of these distributions should be taken as the relevant measure of peer-effect heterogeneity? The raw estimates $\hat{\gamma}_j$ overstate true variation, while the posterior means $\hat{\gamma}_j^*$ understate it because shrinkage pulls values toward the mean to improve prediction accuracy in new contexts. The appropriate measure is therefore the estimated standard deviation of the prior distribution, which provides an unbiased estimate of the true dispersion in peer effects in our setting (Walters 2024). This parameter is informative about the importance of peers in shaping individual academic achievement: One standard deviation of this prior distribution is equivalent to 0.013 SD in grades.

[18] See Walters (2024) for a more detailed discussion.

This implies that being assigned to a section peer whose PVA is one SD above the mean (compared to being assigned an average peer) causes a 0.013 SD increase in grades.

To get a better understanding of the magnitude of this effect, it is helpful to benchmark peer spillovers against better-studied teacher effects. While 0.013 SD appears modest, it is approximately one-third to one-fifth the size of typical teacher value-added estimates. Figure A3 in the appendix compares our PVA estimates to estimates in the literature on the value-added of teachers and school principals. Of the studies summarized in this figure, the median variation for teacher value-added is 0.10 standard deviations, and the median variation for principal value-added is 0.12 standard deviations. This suggests that individual peers, though not as influential as teachers, make a meaningful contribution to performance.[19] Further, based on the distribution of PVA shown in Figure 1 we can infer that replacing a bottom decile peer with a top decile peer improves performance by about 0.046 SD. In Sections 4.2 and 4.3, we aggregate the PVA of all peers at the section level and examine how much the overall peer group matters. We document that modest individual spillovers cumulate into meaningful aggregate achievement differences.

4.2 Cross-Validation of Empirical Bayes PVA Measures

Leave-one-out cross-validation: We examine how well empirical Bayes PVA predicts peer effects outside the estimation sample in a leave-one-out cross-validation. Random assignment of students to peers eliminates selection concerns, so this exercise directly tests the causal link between our empirical Bayes PVA estimates and student performance.

To create the cross-validation out-of-sample exercise, we iterate over all observed courses: each course c' is held out once as a test set, while the remaining courses form the training set are used to fit the model according to equation (6):

---

[19] It is not clear whether teacher value-added and peer value-added are directly comparable, as the appropriate metric depends on the stage of the empirical Bayes procedure one considers. Raw estimates are noisy and therefore not meaningful, while posterior means are shrunken and not suitable for variance comparisons. Estimates of the latent prior distribution provide the most appropriate benchmark for comparing the true underlying variation, but most teacher value-added papers do not report them and instead focus their discussion on the variation in their empirical Base estimates that we report in Figure A3.

$$\hat{u}_{ics}^{(-c')} = \beta_0^{(-c')} + \sum_{j \neq i} \gamma_j^{(-c')} D_{ijcs} + \varepsilon_{ics}^{(-c')}. \qquad (12)$$

Following equation (12), we re-estimate PVA excluding course c' . The residual student grades $\hat{u}_{ics}^{(-c')}$ are regressed on peer dummies D_{ijcs}, producing jackknife leave-out raw PVA estimates $\hat{\gamma}_j^{(-c')}$.[20] Next, we apply empirical Bayes shrinkage using standard errors obtained from a regression based on the full estimation sample.[21] Finally we aggregate the jackknife leave-out-measures of PVA at the section level and compute the sum of PVA by section, mimicking the potential outcome equation of equation (3):

$$\widehat{\Theta}_{ics}^{(-c')} = \sum_{j \in G_{ics}} \hat{\gamma}_j^{(-c')}. \qquad (13)$$

We label this section-level sum of jackknife empirical Bayes PVA the "peer group PVA" from here on. We then regress individual student achievement on this peer group empirical Bayes PVA, controlling for course and student fixed effects and own empirical Bayes PVA:

$$Y_{ics} = \alpha_i + \delta_c + \rho \widehat{\Theta}_{ics}^{(-c')} + \eta \hat{\gamma}_{ics}^{(-c')} + \varepsilon_{ics}. \qquad (14)$$

Using these jackknife leave-out measures of PVA to predict performance avoids mechanical correlation by ensuring that the same estimation errors do not appear on both sides of the regression equation. The coefficient ρ thus may be interpreted as the causal effect of being assigned a peer group with a higher empirical Bayes PVA estimate and serves as a mean forecast-unbiasedness test.

Figure 2 visualizes regression results on the relationship between (section-level) peer group empirical Bayes PVA and student performance. In Panel A, we first

[20] The resulting jackknife raw PVA measure is highly related to our original raw PVA. Regressing raw PVA on jackknife leave-out PVA yields a coefficient of $\hat{\beta} = 0.935$.

[21] Using standard errors obtained from jackknife regressions yields less precise but overall, very similar results.

establish that empirical Bayes PVA is not correlated with students' pre-assignment GPA at baseline. This balancing test confirms the randomness of the peer assignment discussed in Section 2.1.[22] In Panel B we show that peer group empirical Bayes PVA is forecast-unbiased and significantly predicts exam performance at the end of the term. A one unit increase in peer group empirical Bayes PVA predicts a 1.15 SD increase in performance for the randomly assigned students (SE = 0.17).[23] We cannot reject that the validation coefficient is equal to 1. This cross-validation thus establishes that the jackknife measure of PVA is a valid, mean-unbiased, predictor of out-of-sample student performance.[24] It highlights that empirical Bayes PVA captures meaningful variation in peer effects, rather than overfitting noise, and that it isolates a causal contribution of peers to student performance.

Importance of a pool of peers. In Section 4.1, we have described variation in PVA at the individual peer level. However, in many real-world applications, the policy-relevant unit remains the peer group. To quantify the overall importance of a peer group, we aggregate individual PVA at the section level as in the previous section, following the additive framework in equation (13).

In Figure 3, we compare student performance by a students' quartile of the distribution of peer group PVA. The unlucky bottom quartile acts as a reference group. Lucky students who were assigned to a section with peers from the highest quartile of the peer group PVA distribution have, on average, 5.7 percent of a standard deviation higher grades than unlucky students with peer group PVA in the lowest quartile. These results show that individual peer contributions do not cancel out each other and that, after aggregating peer effects at the group level, we can still detect meaningful aggregate effects of being assigned to more-valuable peers.

---

[22] Appendix Table A2 provides an additional randomization check for the validation exercise and establishes that jackknife PVA is unrelated to predetermined student characteristics.

[23] Appendix Table A3 shows corresponding regression results for raw and shrunken PVA.

[24] In Appendix A2, we complement the parameterized significance test with a permutation test based on 1,000 random peer group reassignments. Figure A6 shows the distribution of the 1,000 placebo validation coefficients. The figure highlights that our empirical validation coefficient clearly falls outside the distribution of simulated validation coefficients.

4.3 Peer Value-Added and Observable Characteristics

Correlates of peer value-added. So far, we have described the variance in individual PVA, and have established the predictive power of empirical Bayes PVA measures for out-of-sample social interactions. We now examine which observable peer characteristics predict raw PVA,[25] in other words, we ask who is a good peer. We only observe a limited set of peer demographic variables in our data, including pre-assignment university GPA, gender, age, and nationality. While these characteristics are highly predictive of students' *own* performance, explaining 34 percent of the variation in own achievement, their predictive power for PVA is close to zero.

Table 4 shows that we observe no systematic correlation between raw PVA and student characteristics. *F*-tests of the (joint) significance of peer observable characteristics in predicting raw PVA do not reject the null. Thus, we do not find predictive power for factors that have been popular in the existing literature on peer effects, such as gender and nationality. Maybe most surprisingly, we also do not observe any significant correlation between PVA and predetermined GPA as our best measure for student ability. This suggests that, in our setting, high-achieving students do not appear to have large systematic effects on the performance of others, although we cannot rule out more modest impacts that our estimates may not have the power to detect.[26]

The teacher value-added literature has documented a similar lack of correlation between value-added and observables (e.g., Rivkin et al 2005). The fact that observed characteristics explain little heterogeneity in peer spillovers suggests that traditional models of peer effects miss meaningful spillovers. However, our data only comprises a

[25] When using PVA as a dependent variable, we rely on raw rather than empirical Bayes estimates since shrinkage trades variance for bias—while raw estimates are noisy but unbiased, shrunken estimates introduce systematic bias that contaminates inference (Kane and Staiger, 2008; Walters, 2024). The intuition is straightforward: raw peer fixed effects suffer from classical measurement error, which inflates standard errors but does not bias coefficients. Empirical Bayes shrinkage, by contrast, systematically pulls estimates toward the mean in proportion to their reliability. While this improves prediction when PVA is used as an independent variable, it induces nonclassical measurement error when used as an outcome, leading to biased inference.

[26] Appendix Table A4 shows that these results remain virtually unchanged after excluding "low-mobility" peers, specifically, the 10 percent of peers with the fewest switches.

handful of student characteristics, and we cannot rule out that PVA is systematically related to other characteristics that we do not observe in our data.

4.4 Robustness, Match Effects, and Additional Simulation Evidence

In this section, we first test how accounting for common section-level shocks affects our results. We then discuss how inference methods and attrition affect results. We provide simulation evidence on how section size, reassignment frequency, and section-level shocks affect identification of PVA. Finally, we discuss possible violations of our main identifying assumptions through potential match effects and their consequences for the interpretation of our results. We delegate details of these analyses to Appendix A2.

Robustness analyses: We conduct a series of robustness checks that are organized around three central concerns: Common shocks at the section level, inference methods, and attrition. Appendix A2.1 provides full details of each exercise.

First, we assess whether our results are sensitive to common shocks at the section level. If such shocks are present, they may be mistakenly attributed to individual peers, potentially inflating the estimated variance of peer effects and overstating their true influence. To assess the relevance of this concern, we conduct two complementary exercises. We show that our estimates are robust to the inclusion of instructor, day-of-week fixed effects, and time-of-day fixed effects. Moreover, we estimate PVA using a mixed-effects model allowing us to separate peer and section-level variance components, confirming that *unobserved* classroom shocks do not significantly confound our PVA estimates.

Second, we investigate how different deconvolution approaches and the computation of standard errors affect the estimated variation of empirical Bayes PVA estimates. Standard errors are a key input for the empirical Bayes approach because they directly affect both the estimated prior of PVA and the subsequent shrinkage of individual peer value-added measures. We show that results are sensitive to how standard errors are computed. We highlight that clustering at the right level is essential to obtain mean-unbiased PVA measures. Additionally, we use a bias-corrected variance estimator following Walters (2024) to recover the underlying variance of PVA without imposing parametric assumptions on the mixing distribution and providing robustness to arbitrary heteroskedasticity.

We also conduct two nonparametric permutation exercises to (1) provide a nonparametric test of the importance of peers, (2) understand how noisy PVA measures are and (3) test whether cross-validation fails when it should. For exercise one, we randomly reassign students within courses and re-estimate our peer-value added measures. We show that individual estimated PVA based on placebo assignments contain significant variation but less than the PVA estimated on the true assignment.[27] In exercise two, we use our cross-validation framework to verify whether peer group PVA has predictive power in placebo assignments. The validation coefficients for placebo assignments are distributed around 0, never exceeding the true validation coefficient.

Third, we provide evidence on how attrition affects the observed distribution of PVA. One concern is that with fewer observations per peer, raw fixed-effect estimates become noisier, widening the empirical distribution. It would also be important to know whether early leavers happen to have systematically different (true) PVA, implying that dropping them could alter the estimation of the latent distribution. We show PVA distributions for "leavers" – students who disappear from the institution— is similar to those who stay. These results suggest that selective attrition does not seem to create meaningful distortions in the observed distribution of PVA.

Simulation evidence: To evaluate the reliability of our estimation approach and to substantiate the core identification strategy, we complement our empirical analysis with Monte Carlo simulations. We delegate the details of these exercises to Section A2.2 in the appendix.

These simulations address two critical questions: First, we examine how far our sample size and data structure supports credible identification of PVA despite limited rounds of reassignment. Based on a known data-generating process (DGP), including a combination of peer effects and i.i.d. noise and varying peer group size and rounds of

[27] The difference between the PVA SD estimated using the true and placebo assignments is quantitatively similar to our reliability ratio of the empirical Bayes approach.

reassignment, we show that estimated hyperparameters converge fast with an increasing number of reassignments, and under the assumed DGP, our empirical data is well suited to estimate the PVA variance reliably. We further show that under the empirical Bayes methodology, hyperparameter estimates converge reliably to the true PVA variance even with large section sizes.

Second, we assess in how far mixed-effects models accurately separate peer and section-level variance components in the presence of unobserved classroom-level shocks. We simulate outcomes based on our actual empirical data structure keeping the true PVA variance constant while varying the relative contribution of section-level noise. As section-level noise increases, the peer variance estimated via OLS becomes increasingly inflated, misattributing section-level variation to individual peer effects. In contrast, mixed-effects models continue to recover the true peer variance reliably, even under substantial section-level noise. We thus interpret the similarity of OLS and mixed-model estimates in our empirical data as evidence for the absence of larger unobserved section-level heterogeneity affecting our results.

Match effects: Match effects arise when the influence of a peer depends on the student they interact with. The existence of such match effects—for example, along shared nationality, gender, or ability—would violate our constant-effects assumption. Consequently, our estimates may understate the true variance in peer effects. Our constant-effects framework recovers only each peer's average contribution across matches and omits the within-peer variation across recipients.

In Appendix A2.3, we estimate an alternative specification that allows PVA to differ between a peers' own group and others. We define the groups along the dimensions of gender, ability, and nationality. A simulation exercise shows that the true variance can, in principle, be recovered by adjusting the estimation equation (6) to allow for match-specific PVA. Yet, this specification comes at the cost of effectively doubling the parameters to be estimated and substantially reducing precision, thus increasing the risk that peer dummies capture noise rather than systematic heterogeneity. Comparing empirically observed match-specific distributions to simulated results indicates no systematic match effects along these dimensions.

We provide more evidence on match effects by examining residuals of equation (6) within cells defined by deciles of students' own ability and peer group PVA following Card, Heining and Kline (2013). Results indicate that, for the bulk of the sample, homogeneous residuals support identification assumptions of constant effects, with some deviations among the tails. Taken together, our empirical evidence suggests an absence of large match effects. However, our limited statistical power means that moderate, match-specific heterogeneity could still go undetected.

4.5 Mechanisms—Evidence from Student Course Evaluations

Observable peer characteristics do not explain PVA. However, students may nevertheless notice and respond to high-value peers through their classroom behavior. To explore possible mechanisms, we analyze individual-level student course evaluations. These evaluations capture perceptions of peer-to-peer learning, group functioning, study effort, and overall course quality. To test these potential channels, we use the same empirical strategy as in the out-of-sample validation in Section 4.2, regressing peer group PVA on individual evaluation responses.

Table 5 shows the estimation results and highlights that being assigned to section peers with a higher PVA leads to a better learning environment. Students exposed to higher value-added peers report improved group functioning and learning more from their peers. A one standard deviation increase in peers' PVA improves the "Working in a study group with peers helped me to better understand course material" item by 0.016 standard deviations and improves perceived group functioning captured by the item "My tutorial group has functioned well" by 0.021 standard deviations. In addition, students with better peers also report increased self-study hours. This suggests that valuable peers act as a complement and not as a substitute for studying alone. Finally, better peers also lead to an increased overall course evaluation, as captured by responses to the item "Please give an overall grade for the quality of this course" by 0.027 standard deviations.

Taken together these results highlight that, while peers' observable characteristics are poor predictors of PVA, fellow students notice and respond to the presence of high value-added peers in their section. Together with the results from the out-of-sample validation exercise, these estimates confirm that peer value-added captures noticeable, systematic spillovers and not merely overfitting random noise.

5. Discussion

5.1 Relationship with Prior Work on Peer Effects

PVA uncovers social spillovers that conventional linear-in-means (LIM) specifications simply miss. We find substantial dispersion in PVA, yet this variation is only weakly correlated with standard student observables such as prior GPA, gender, or nationality. Hence, most of the ways in which classmates help or hinder one another operate through channels that LIM models—whose regressors are peer averages of observables—cannot detect.

Appendix A3 formalizes the link between PVA and LIM models: the LIM coefficient equals the covariance between PVA and the chosen observable, scaled by 1/(n – 1), where *n* is peer group size. If that covariance is small, LIM will declare "no peer effect" even when large unobserved spillovers are present. Put differently, LIM detects peer influences only when they happen to line up with the particular observable in the regression. This is a blind spot that PVA overcomes by estimating each student's total contribution to classmates' learning directly, regardless of what observed characteristics valuable students have.

Our work complements some previous studies that have acknowledged the potential role of unobservable peer characteristics without estimating peer value-added. Fruehwirth (2014) argues that peer achievement itself is unlikely to produce spillovers but that it can act as a useful proxy for unobservable sources of peer spillovers, such as ability, motivation, or effort. Following a similar rationale, Arcidiacono et al. (2012) model spillovers as linear combinations of student fixed effects to capture peers' unobserved abilities.[28] Closest to our paper, three recent studies estimate individual-

[28] Note that this approach is significantly different from our strategy. Arcidiacono et al. (2012) model peer spillovers as a function of unobservable peer ability, proxied by peer fixed effects from a regression of peer *j's* performance on peer *j* fixed effects. In contrast, our PVA measures capture the expected spillover of peer *j* on the performance of student *i*. This spillover does not necessarily have to work through peer *j*'s ability. In fact, our results do not show a strong relationship between student ability and PVA.

specific contributions to team production. These papers, however, do not use individual performance as an outcome. Arcidiacono et al. (2017) estimate player-level contributions to team performance based on National Basketball Association data. However, given their lack of information on individual player productivity, they cannot rely on direct player–peer pair observations, and they face challenges related to the selection of players into games. The authors solve these issues by relying on a saturated structural model of player productivity, recovering estimates of player-specific spillovers. Weidmann and Deming (2021) conduct a lab experiment in which players solve tasks in teams. The authors show that some players consistently contribute more to joined team production than others. These players are labeled *team players.* Team players do not differ in terms of IQ or personality, but score highly on a measure of social intelligence. Bonhomme (2021) estimates AKM models for network data of economists and inventors moving between faculties/teams to separate individual contributions to team production. We complement these papers by focusing on *individual* instead of *team* production output. Thus, our paper speaks to situations in which peers affect individual performance, as is typical for educational settings.

5.2 Policy Relevance, Risks and Limitations

Policy Relevance: Peer value-added measures could become policy relevant in a variety of education and workplace settings by allowing policy makers to assign individual peers based on their ability to improve others' outcomes. For example, valuable peers could be strategically assigned to provide targeted peer support for disadvantaged students. PVA could be used to identify promising mentors and teaching assistants in an educational environment. They could also inform admission policies for selective institutions. Admitting more valuable peers could increase the collective performance within a cohort. In firms, managers could use peer value-added measures to identify and reward individuals who increase team productivity—not solely through their direct contributions but also through their positive impact on colleagues. This could be especially relevant for large organizations with frequent variation in team composition.

Potential Risks and Limitations: While we demonstrate that PVA is a forecast-unbiased predictor for out-of-sample spillovers, the noise-to-signal ratio is low. Our average shrinkage factor suggests individual level PVA measures contain only a 5–7 percent signal. This points to some risks and limitations in using PVA in practice. First, using these noisy individual estimates may lead to considerable misclassification, that is, incorrectly labeling students as high or low value-added. Second, even if individual peer value-added measures were estimated without noise, reassignment based on PVA could unintentionally disrupt existing learning dynamics. As documented by Carrell et al. (2013), well-intended reassignment policies can lead to adverse outcomes. Similarly, reallocating students based on PVA could undermine classroom cohesion or trigger unintended behavioral responses. Peer effects following nonrandom reassignment may not be linear or additive. Students who are deemed to be particularly valuable peers might, for example, exert disproportionately smaller—or larger—effects when assigned to sections with the weakest students of a cohort. If policies assume linear peer effects, they risk misjudging the true consequences of peer reallocation. Taken together, these concerns suggest we should be careful when using empirical Bayes PVA measures for real-world assignment policies.

6. Conclusion

We introduce a new way to quantify the importance of peers. Building upon methods developed in the teacher effectiveness literature, we conceptualize and empirically estimate the value-added of an individual peer. In contrast to existing approaches to estimate peer effects, peer value-added captures the total contribution to the performance of others, capturing both observable and unobservable characteristics that create spillovers. The PVA approach does not require taking a stand about which peer observables create spillovers.

By leveraging repeated randomization of students into small teaching sections, we estimate PVA as random effects in an empirical Bayes framework and recover the distribution of peer value-added**.** We find significant variation in how much peers affect university performance and establish that peer value-added is a valid predictor of spillovers in new social interactions. We further quantify the importance of an overall pool of peers. A student in a top quartile peer group has, on average, 5.7 percent of a standard deviation higher performance than a student with a peer group in the bottom quartile of the distribution.

PVA is largely unrelated to peers' observable characteristics. Most notably, it is not correlated with past achievement, implying that higher performers do not necessarily bring out higher performance in others, which is an implicit assumption in many models of ability peer effects. While observed student characteristics do not predict PVA, students notice and react to more-valuable peers. Having more-valuable peers improves classroom climate and student effort. Their presence strengthens group functioning and the quality of peer-to-peer interactions as well as raises students' self-study hours.

Our conceptual framework and estimation provide a new perspective on how to think about peer effects. Peer value-added allows for estimating peer effects without priors about which observable peer characteristics generate spillovers. Further, it can be estimated in the absence of information on observable peer characteristics. The PVA approach proposed in this paper can be applied in any setting with repeated observations of individual performance. This approach may therefore help to (re-)assess the importance of peers in a variety of educational and work settings.

REFERENCES

Aaronson, D., Barrow, L., and Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, *25*(1), 95–135.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica* 67(2), 251–333.

Antonovics, K., Black, S. E., Cullen, J. B., and Meiselman, A. Y. (2022). Patterns, determinants, and consequences of ability tracking: Evidence from Texas public schools (No. w30370). National Bureau of Economic Research.

Arcidiacono, P., Foster, G., Goodpaster, N., and Kinsler, J. (2012). Estimating spillovers using panel data, with an application to the classroom. *Quantitative Economics,* 3, 421–470.

Arcidiacono, P., Kinsler, J., and Price, J. (2017). Productivity spillovers in team production: Evidence from professional basketball. *Journal of Labor Economics*, 35(1), 191–225.

Bacher-Hicks, A., & Koedel, C. (2023). Estimation and interpretation of teacher value added in research applications. *Handbook of the Economics of Education*, *6*, 93–134.

Bartanen, B., Husain, A. N., and Liebowitz, D. D. (2024). Rethinking principal effects on student outcomes. *Journal of Public Economics*, *234*, 105115.

Bau, N., & Das, J. (2020). Teacher value added in a low-income country. *American Economic Journal: Economic Policy*, *12*(1), 62–96.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy*, *75*(4, Part 1), 352–365.

Bonhomme, S. (2021). Teams: Heterogeneity, sorting, and complementarity. Forthcoming in *Proceedings of the 15th World Congress of the Econometric Society.*

Booij, A. S., Leuven, E., and Oosterbeek, H. (2017). Ability peer effects in university: Evidence from a randomized experiment. *Review of Economic Studies*, 84(2), 547–578.

Branch, G. F., Hanushek, E. A., and Rivkin, S. G. (2012). Estimating the effect of leaders on public sector productivity: The case of school principals (No. w17803). National Bureau of Economic Research.

Burke, M. A., & Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, *31*(1), 51–82.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, *29*(2), 238–249.

Card, D., Heining, J., and Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality. *Quarterly Journal of Economics,* 128(3), 967–1015.

Carrell, S. E., Fullerton, R. L., and West, J. E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, *27*(3), 439–464.

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

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2593–2632.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–2679.

Chiang, H., Lipscomb, S., and Gill, B. (2016). Is school value added indicative of principal quality?. *Education Finance and Policy*, *11*(3), 283–309.

Dhuey, E., and Smith, J. (2018). How School Principals Influence Student Learning. *Empirical Economics*, 54(2), 851–882.

Elsner, B., Isphording, I. E., and Zölitz, U. (2021). Achievement rank affects performance and major choices in college. *The Economic Journal,* 131(640), 3182-3206.

Epple, D., and Romano, R. (2011). Peer effects in education: A survey of the theory and evidence. *Handbook of Social Economics,* 1(1), 1053–1163.

Feld, J., and Zölitz, U. (2017). Understanding peer effects – On the nature, estimation, and channels of peer effects. *Journal of Labor Economics*, 35(2), 387–428.

Feld, J., Salamanca, N., and Zölitz, U. (2018) Are professors worth it? The value-added and costs of tutorial instructors. *Journal of Human Resources*, 0417.

Fruehwirth, J. C. (2014). Can achievement peer effect estimates inform policy? A view from inside the black box. *The Review of Economics and Statistics,* 96(3), 514–523.

Garlick, R. (2018). Academic peer effects with different group assignment policies: Residential tracking versus random assignment. *American Economic Journal: Applied Economics*, 10(3), 345–369.

Gilraine, M., Gu, J., & McMillan, R. (2020). A new method for estimating teacher value-added (No. w27094). National Bureau of Economic Research.

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of human Resources*, 351–388.

Hanushek, E. A., and Rivkin, S. G. (2006). Teacher quality. *Handbook of the Economics of Education*. 1051–1078.

Jacob, B. A., and Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101–136.

Kane, T. J., Rockoff, J. E., and Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*(6), 615–631.

Kane, T. J., and Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. *NBER Working Paper 14607*.

Kline, P., Saggio, R., and Sølvsten, M. (2020). Leave-out estimation of variance components. *Econometrica*, *88*(5), 1859–1898.

Koedel, C., and Betts, J. R. (2011). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education Finance and Policy*, *6*(1), 18–42.

Koedel, C., Mihaly, K., and Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review,* 47, 180–195.

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

Nye, B., Konstantopoulos, S., and Hedges, L. V. (2004). How Large are Teacher Effects? *Educational Evaluation and Policy Analysis*, 26(3), 237–257.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, *73*(2), 417–458.

Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review*, 94(2), 247–252.

Rose, E. K., Schellenberg, J. T., and Shem-Tov, Y. (2022). The effects of teacher quality on adult criminal justice contact (No. w30274). National Bureau of Economic Research.

Rothstein, J. (2010). Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement. *Quarterly Journal of Economics*, 125(1), 175–214.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? in E. Hanushek, S. Machin, L. Woessmann (eds.), *Handbook of the Economics of Education*, Elsevier, edition 1, volume 3, number 3, 249–277.

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: Two steps forward? *Annual Review of Economics*, 6(1), 253–272.

Todd, P. E., and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal,* 113(485), F3–F33.

Walters, C. (2024). Empirical Bayes methods in labor economics. In *Handbook of Labor Economics* (Vol. 5, pp. 183-260). Elsevier.

Wang, L. C. (2010). Peer effects in the classroom: Evidence from a natural experiment in Malaysia. *Unpublished manuscript*.

Weidmann, B., and Deming, D. J. (2021). Team players: How social skills improve team performance. *Econometrica*, 89(6), 2637–2657.

TABLES AND FIGURES

Table 1: Descriptive Statistics

	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
A. Student Demographic Variables					
Age	3,631	19.99	1.72	16.33	33.92
Female	3,970	0.38	0.49	0.00	1.00
Dutch	3,970	0.24	0.43	0.00	1.00
German	3,970	0.47	0.50	0.00	1.00
B. Student Performance					
Pre-assignment GPA	46,913	6.89	1.14	2.25	10.00
Course grade	53,569	6.57	1.82	1.00	10.00
C. Variation in Course-Year Observations					
Sections per course-year observation	360	12.35	12.01	1.00	43.00
Students per course-year observation	360	148.80	156.69	5.00	583.00
D. Variation in Sections					
Students per section	4,447	12.05	2.01	3.00	16.00
Share of identified peers per section	4,447	0.89	0.11	0.50	1.00
E. Variation in Students					
Number of individual students	3,970				
Number of individual peers	2,692				
Identified peers per student	3,970	138.07	75.32	15	299
Repeated student–peer interactions	1.34				

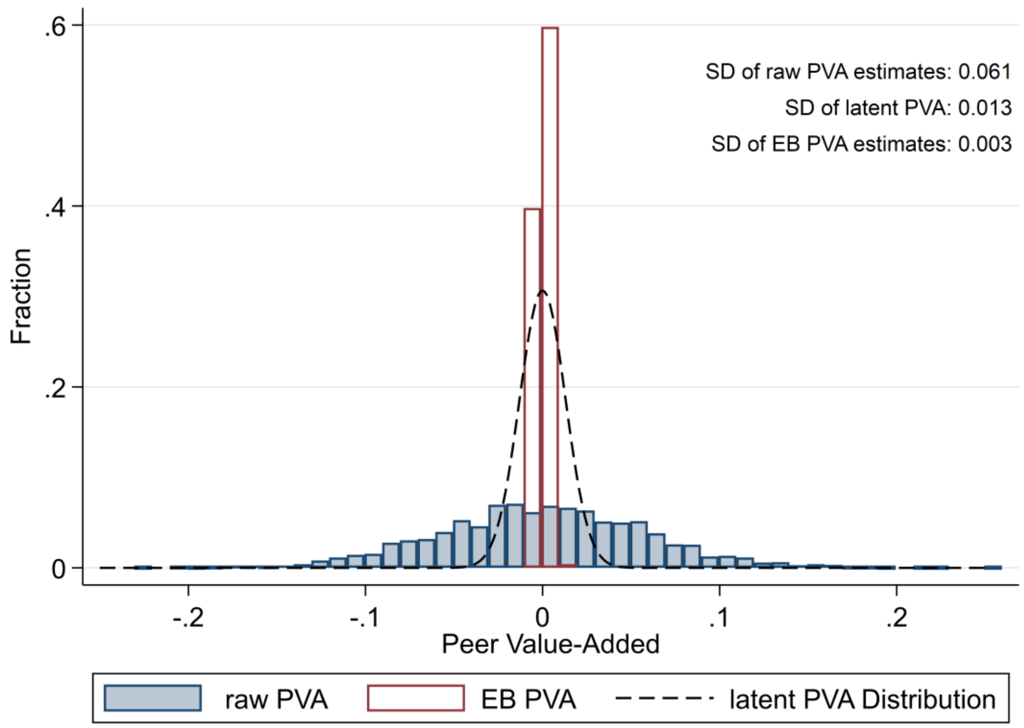
Note: This table shows descriptive statistics of the estimation sample. "SD" refers to the standard deviation of the respective variable. Panel A reports individual student demographic characteristics. Panel B reports measures of student performance: pre-section assignment grade point average (GPA) and end-of-course grades at the student-course level. Pre-assignment GPA in Panel B is missing for students in the first period. Panels C, D, and E report on the variation of course-year, sections, and students. Students are classified as an *identified peer* if they are eligible for peer value-added estimation. The share of identified peers is smaller than one because students must be in at least eight different course-year observations and belong to the two main study programs.

Table 2: Test for Random Assignment of Students to Sections

	(1)	(2)			(3)		
Dependent variable:	Number of total tests performed	Number significant at:			Percent significant at:		
		5%	1%	0.1%	5%	1%	0.1%
Female	337	13	2	0	3.9%	0.6%	0.0%
GPA	326	32	6	1	9.8%	1.8%	0.3%
Age	337	11	1	0	3.3%	0.3%	0.0%
ID rank	337	19	2	0	5.6%	0.6%	0.0%
Dutch	335	6	2	0	1.8%	0.6%	0.0%
German	337	5	2	0	1.5%	0.6%	0.0%
Total	2009	86	15	1	4.28%	0.75%	0.05%

Note: This table is based on separate OLS regressions with gender, GPA, age, and ID rank (a proxy for student tenure at the university) as dependent variables. The explanatory variables are a set of section dummies. Columns (1) and (2) show in how many regressions the *F*-statistic of joint significance of all included section dummies is statistically significant at the 5 percent, 1 percent, and 0.1 percent levels, respectively. Differences in the number of tests performed (column 3) are due to missing observations for some of the dependent variables. German and Dutch are mechanically balanced due to the stratified assignment by nationality. Figure A2 shows histograms of the *p*-values for each performed test from the underlying regressions.

Figure 1: The Distribution of Peer Value-Added



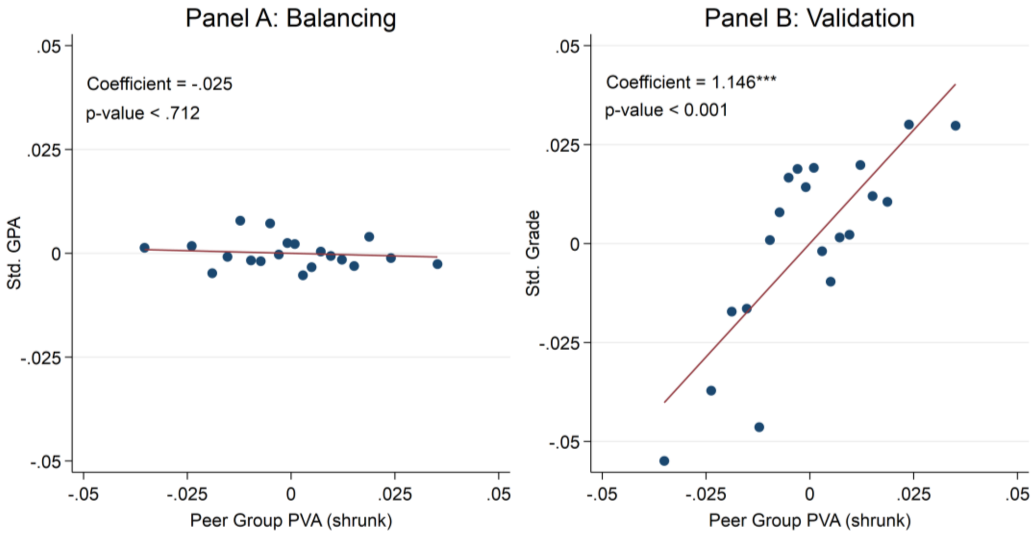
Note: This figure shows the value-added estimates at the level of 2,692 peers. Underlying grades are standardized to a mean of zero and unit variance. The blue bars display the spread in raw PVA estimated in equation (6). The hollow red bars display the spread of the shrunken empirical Bayes PVA estimates. The dashed line indicates the estimated distribution of latent PVA (mixing distribution) based on the estimated hyperparameters as shown in equations (9) and (10).

Table 3: Test of Joint Significance of Peer Dummies

Panel A: All	(1) *F* Stat	(2) *p*-value	(3) Clustering	(4) Number of Restrictions
	5.97	< 0.001	Student Level	2,692
	6.57	< 0.001	Section Level	2,692
Panel B: Peer Partitioning	***F* Stat**	***p*-value**	**Clustering**	**Number of Restrictions**
Block 1	2.71	< 0.001	Student + Course Level	337
Block 2	2.54	< 0.001	Student + Course Level	337
Block 3	2.32	< 0.001	Student + Course Level	337
Block 4	3.28	< 0.001	Student + Course Level	337
Block 5	2.84	< 0.001	Student + Course Level	336
Block 6	2.43	< 0.001	Student + Course Level	336
Block 7	3.03	< 0.001	Student + Course Level	336
Block 8	3.24	< 0.001	Student + Course Level	336
Panel C: Random Draws	**Avg. *F* Stat**	**Avg. *p*-value**	**Clustering**	**Number of Restrictions**
	2.97	< 0.001	Student + Course Level	359

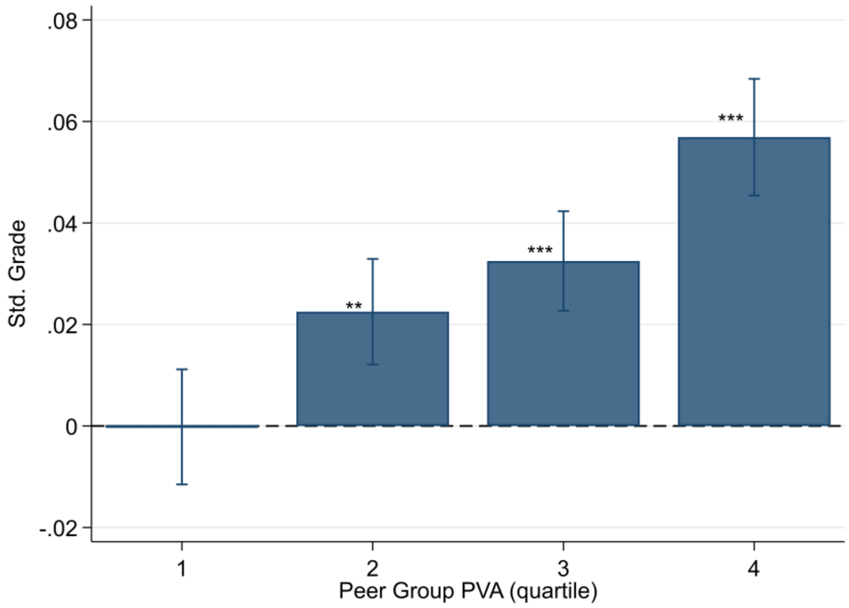
Note: This table shows tests of joint significance for the peer dummy variables. Panel A reports the *F*-statistic and *p*-value when using all peer dummies while clustering on the student or section level. Panel B reports the *F*-statistic and *p*-value when splitting peer dummies into individual and unique separate blocks and using two-way clustering on the student and course level. Peer dummies are split by first sorting all peers by their student ID and then repeatedly assigning the values one to eight throughout the list. Panel C reports the average *F*-statistic and *p*-value for 1,000 x 359 randomly drawn peer dummies. Column (4) shows the number of jointly tested peer dummies.

Figure 2: Balancing and Jackknife Validation of Peer Value-Added



Note: This binned scatterplot visualizes the relationship between student GPA (Panel A), student end-of-course grades (Panel B) and Peer Group PVA, which is constructed based on a course-leave-out sample that excludes the course in which we observe the grade we use as a dependent variable. Peer Group PVA sums up the estimate empirical Bayes PVAs of all peers within the same section. The point estimates and significance levels are obtained from a regression of standardized GPA (Panel A) or standardized grade (Panel B) on Peer Group PVA, and conditioning on own empirical Bayes PVA, student-level fixed effects, and course-level fixed effects. One observation is one student-section observation. N = 55,569. The R^2 of the validation regression is 0.61. All *p*-values are based on two-way clustering at the student and course level. *** *p*<0.01, ** *p*<0.05, * *p*<0.1.

Figure 3: Aggregate Variation in Peer Value-Added Across All Peers



Note: This figure shows the effect of Peer Group PVA on standardized end-of-course grades. Peer Group PVA is constructed based on a course-leave-out sample that excludes the course in which we observe the grade we use as a dependent variable. Peer Group PVA sums up the empirical Bayes PVA estimates of all peers within the same section. We categorize Peer Group PVA into quartiles at the course level after dropping students for which we do not estimate PVA and students with the lowest 10% of reassignments. We control for the individual empirical Bayes PVA estimate, student-level fixed effects, and course-level fixed effects. Standard errors are clustered at the student and course level. The stars represent the significance of testing the quartiles against the benchmark quartile Q1. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Error bars indicate the 95% confidence interval.

Table 4: Correlates of Peer Value-Added

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dep. Var.: raw PVA						
Female	0.001						0.001
	(0.002)						(0.002)
Mean GPA (std.)		0.000					-0.000
		(0.001)					(0.001)
Age (at Enrollment)			0.000				0.000
			(0.001)				(0.001)
Study Cohort = 2010				-0.000			-0.001
				(0.004)			(0.004)
Study Cohort = 2011				-0.001			-0.001
				(0.003)			(0.003)
Study Cohort = 2012				-0.001			-0.001
				(0.003)			(0.003)
Dutch					-0.003		-0.002
					(0.003)		(0.004)
German						0.002	0.000
						(0.002)	(0.004)
Observations	2,692	2,692	2,692	2,692	2,692	2,692	2,692
Adjusted R-squared	-0.000	-0.000	-0.000	-0.001	-0.000	-0.000	-0.002

Note: This table reports OLS regression results of raw PVA on peer observable characteristics with one observation per peer. GPA is standardized to have a mean of zero and standard deviation of one for our estimation sample. The baseline category for the study cohort is 2009. We use robust standard errors. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Impact of Peers on Student Course Evaluations

Panel A: Raw Values	(1) Working in study group with peers helped me to better understand course material	(2) Tutorial group has functioned well	(3) Weekly study hours (average)	(4) Perceived overall course quality
Std. Peer Group PVA (raw)	0.016*	0.021**	0.168**	0.027***
	(0.009)	(0.009)	(0.067)	(0.009)
Observations	18,718	18,814	17,429	19,859
Adjusted R-squared	0.208	0.200	0.487	0.299
Own PVA	Yes	Yes	Yes	Yes
Course-Time FE	Yes	Yes	Yes	Yes
Student FE	Yes	Yes	Yes	Yes

Panel B: Shrunken Values	Working in study group with peers helped me to better understand course material	Tutorial group has functioned well	Weekly study hours (average)	Perceived overall course quality
Std. Peer Group PVA (EB)	0.014	0.023**	0.126*	0.029***
	(0.009)	(0.009)	(0.065)	(0.008)
Observations	18,718	18,814	17,429	19,859
Adjusted R-squared	0.208	0.201	0.487	0.299
Own PVA	Yes	Yes	Yes	Yes
Course-Time FE	Yes	Yes	Yes	Yes
Student FE	Yes	Yes	Yes	Yes

Note: This table shows the relationship between course evaluation outcomes and standardized Peer Group PVA, which is constructed based on a course-leave-out sample that excludes the course in which we observe the grade we use as a dependent variable. Peer Group PVA sums up the estimated PVAs of all peers within the same section. Panel A reports the results for standardizing Peer Group PVA based on raw PVA estimates. Panel B reports the results for standardizing Peer Group PVA based on empirical Bayes PVA estimates. The items of the evaluation are: "Working in tutorial groups with my fellow students helped me to better understand the subject matters of this course," "My tutorial group has functioned well," "How many hours per week on the average did you spend on self-study," and "Please give an overall grade for the quality of the course." We control for the individual shrunken PVA estimate, student-level fixed effects, and course-level fixed effects. Standard errors are clustered at the student and course level. *** p<0.01, ** p<0.05, * p<0.1.

ONLINE APPENDIX

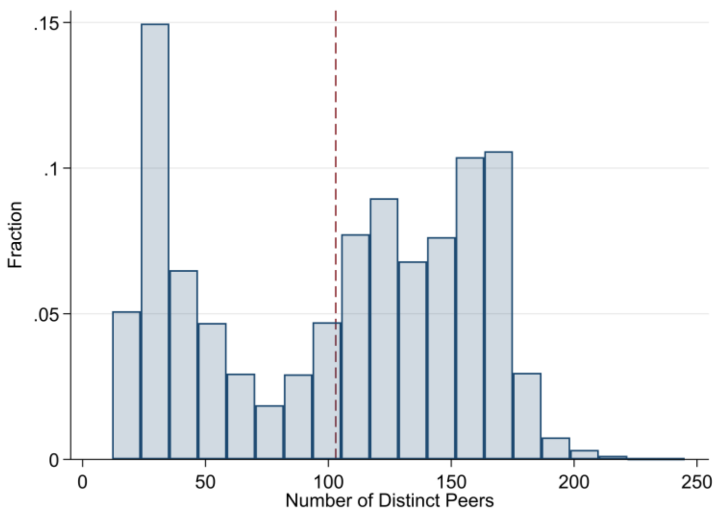
APPENDIX A1: Additional Tables and Figures

Table A1: Connectedness of Students

How are students connected?	(1) % Connected	(2) N
Direct link	3.80%	3,970
Via 1 link	55.30%	3,970
Via 2 links	99.20%	3,970
Via 3 links	100.00%	3,970

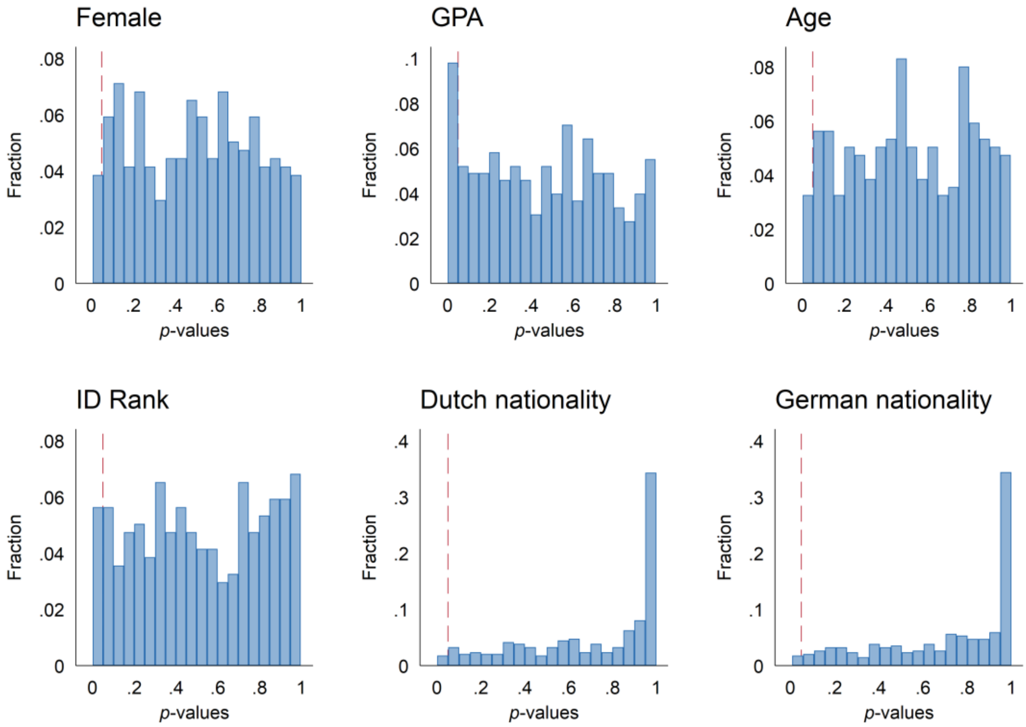
Note: This table shows how many uniquely encountered peers are interacting with each student directly or via links. Direct links refer to being in the same section as the peer.

Figure A1: Number of Distinct Peers



Note: This figure shows the distribution of the number of unique peers that each student meets. We only count peers for which we estimate PVA. The dotted red line indicates the average. n = 3,970.

Figure A2: Randomization Check—Distribution of *p*-values



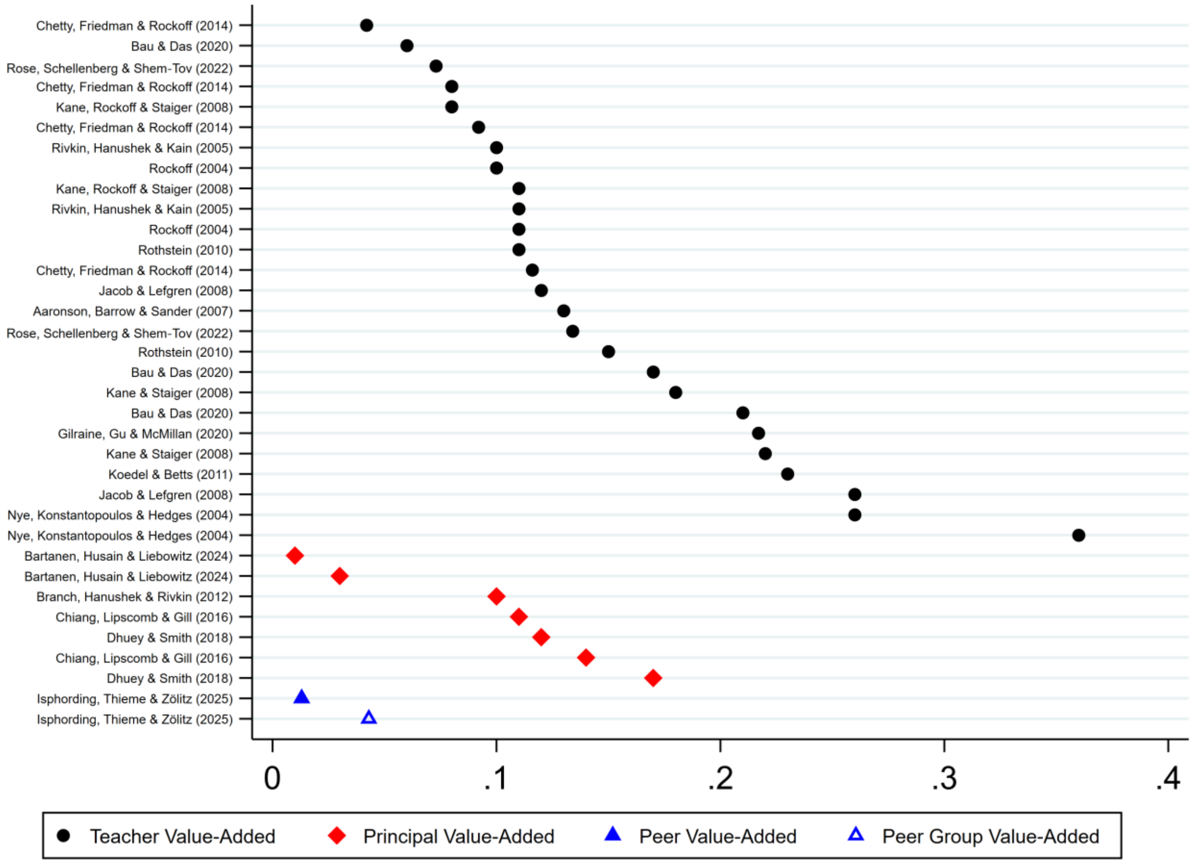
Note: These histograms show *p*-values from all the regressions reported in Table 2. The vertical line in each histogram shows the 5 percent significance level. Dutch and German are mechanically "over-balanced" due to the stratified assignment based on nationality.

Table A2: Randomization Check for Leave-out Jackknife Validation

	(1) Female	(2) Age	(3) GPA	(4) Dutch	(5) German
Peer Group PVA	0.044	0.134	0.124	0.089	-0.169*
	(0.123)	(0.315)	(0.282)	(0.078)	(0.087)
Observations	53,569	52,356	46,913	53,569	53,569
Adjusted R-squared	0.019	0.229	0.063	0.041	0.050
Own PVA	Yes	Yes	Yes	Yes	Yes
Course-Time FE	Yes	Yes	Yes	Yes	Yes

Note: This table reports the results of regressions of baseline characteristics on Peer Group PVA, which is constructed based on a course-leave-out sample that excludes the course in which we observe the grade we use as a dependent variable. Peer Group PVA sums up the estimated empirical Bayes PVAs of all peers within the same section. We control for the individual empirical Bayes PVA estimate and course-level fixed effects. Standard errors are clustered at the student and course level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Figure A3: Comparison of Teacher, Principal, and Peer Value-Added



Note: This figure summarizes the variation in value-added estimates of teachers (black circles), principals (red diamonds), peers (blue triangle), and peer groups (hollow blue triangle) in recently published studies. All estimates refer to standard deviations of value-added distributions constructed for standardized performance measures.

Table A3: Validation of PVA Measures

	(1)	(2)
	Std. Grade	Std. Grade
Peer Group PVA (raw)	0.071***	
	(0.022)	
Peer Group PVA (EB)		1.146***
		(0.177)
Observations	53,569	53,569
Adjusted *R*-squared	0.573	0.573
Own PVA	Yes	Yes
Course-Time FE	Yes	Yes
Student FE	Yes	Yes

Note: This table shows the relationship between standardized end-of-course grades and Peer Group PVA, which is constructed based on a course-leave-out sample that excludes the course in which we observe the grade we use as a dependent variable. Peer Group PVA sums up the estimated PVAs of all peers within the same section. Column (1) shows the result when using the raw PVA values. Column (2) shows the result when using the empirical Bayes PVA values. We control for the individual PVA estimate, student-level fixed effects, and course-level fixed effects. Standard Errors are clustered at the student and course level. *** p<0.01, ** p<0.05, * p<0.1.

Table A4: Robustness of Table 4—Dropping 10% of Students with Fewest Section Reassignments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dep. Var.: raw PVA						
Female	0.000						0.0010
	(0.002)						(0.002)
Mean GPA (std.)		-0.001					-0.0020
		(0.001)					(0.001)
Age (at Enrollment)			0.000				0.0000
			(0.001)				(0.001)
Study Cohort				0.000			0.0000
				(0.001)			(0.001)
Dutch					-0.003		-0.0020
					(0.003)		(0.004)
German						0.002	0.0020
						(0.002)	(0.004)
Observations	2,429	2,429	2,429	2,429	2,429	2,429	2,429
Adjusted *R*-squared	0.000	0.000	0.000	0.000	0.000	0.000	-0.001

Note: This table shows the relationship between individual characteristics and estimated raw PVA for a sample in which we drop the 10% of students with the fewest section reassignments. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

APPENDIX A2: *Robustness Checks, Match Effects, and Additional Simulation Evidence*

A2.1 Robustness Checks

Robustness analyses: We conduct a series of empirical exercises to assess the robustness of our peer-value added (PVA) estimates. These exercises are organized around three central concerns.

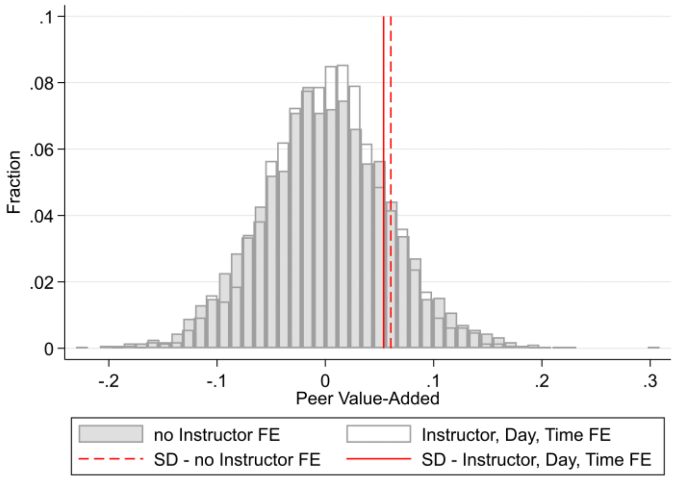
First, we examine whether section-level influences affect our estimates of peer effects and potentially bias the estimated variance. Second, we assess whether the estimated standard deviation of PVA is appropriately recovered. Third, we investigate whether selective attrition systematically distorts the distribution of PVA.

Robustness to section-level influences: Besides peer effects, other section-specific factors such as the instructor, time of day, day of the week, or other unobserved factors may influence individual student outcomes. Despite random assignment, such section-level shocks may affect our peer estimates. In particular, they could inflate the estimated variance of peer effects, making peers appear more influential than they truly are. To assess the relevance of these concerns, we conduct two complementary exercises.

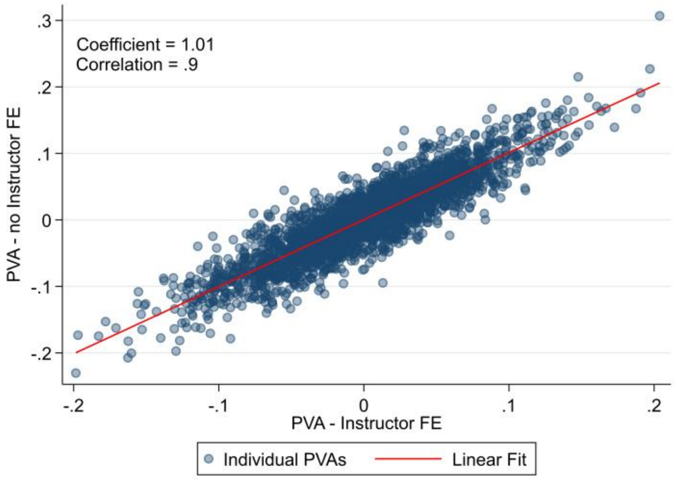
First, we control for observable section-level characteristics before estimating PVA. This means that in our residualizing procedure we additionally include instructor, time of day, and day of the week FE to remove all performance variation that is attributable to these factors. Figure A4 compares the PVA distributions of our main specification and this section-level–adjusted specification. The distributions are highly overlapping and individual estimates exhibit a strong correlation of about 0.9. Additionally, Table A5 shows that our main estimates are robust to the inclusion of instructor, day-of-week fixed effects, and time-of-day fixed effects. Table A5 reports the estimated standard deviation of PVA for different estimation specifications in which student performance is residualized before PVA construction additionally based on instructor-level, and day and time of section level fixed effects. Adding instructor fixed effects decreases the estimated raw variance by about 10 percent. Additionally, controlling for day and time of section level FE has no additional effect on the estimated variance.

Figure A4: Raw Peer Value-Added vs. Peer Value-Added with Instructor, Day, and Time Fixed Effects

Panel A: Variation in PVA with Different Set of Fixed Effects



Panel B: PVA vs PVA net of Observable Section Shocks



Note: Panel A shows the distribution of raw PVA estimated with and without instructor-level, day-level, and time-level fixed effects. The dashed red line indicates the standard deviation using raw PVA estimates and without instructor-level, day-level, and time-level fixed effects. The solid red line indicates the standard deviation of estimates with instructor-level, day-level, and time-level fixed effects. Panel B shows the peer-level relationship between raw PVA and raw PVA net of instructor-level, day-level, and time-level fixed effects. The correlation between both PVA estimates is 0.90. A linear regression using the raw PVA net of instructor-level, day-level, and time-level fixed effects as the explanatory variable yields a coefficient of 1.01 and an R^2 of 0.81.

Table A5: Variation in Peer Value-Added

Specification	(1) SD of raw PVA estimates	(2) SD of latent PVA	(3) Correlation with PVA (Main specification)
Main specification	0.061	0.013	
Instructor FE in 1st Step	0.054	-	0.899
Instructor, Day + Time FE in 1st Step	0.054	-	0.896

Note: This table reports the estimated standard deviation of PVA for different estimation specifications. For the main specification, we residualize standardized end-of-course grades with student-level and course-level fixed effects and estimate PVA with the obtained residuals on a set of dummy variables, clustering standard errors on the student and course level. In the alternative specifications, we increase the set of fixed effects that are used for residualizing standardized end-of-course grades. We use student-level, course-level, and instructor-level fixed effects or student-level, course-level, instructor-level, and day and time of section-level fixed effects. Column (1) reports the estimated standard deviation using the raw estimates. Column (2) reports the estimated standard deviation of the estimated latent distribution using empirical Bayes. Column (3) shows the correlation of the individual raw PVA estimates from the main alternative specifications.

Table A6: Mixed Model

Specification	(1) SD of raw PVA estimates	(2) Correlation with PVA (Main specification)
Main specification	0.061	-
Mixed-model specification	0.061	1.00

Note: This table reports the standard deviation of raw PVA estimates for our main specification and an alternative mixed-model specification. For the main specification, we residualize standardized end-of-course grades with student-level and course-level fixed effects and estimate PVA with the obtained residuals on a set of dummy variables. The mixed specification fits a linear mixed model with the same peer dummies as fixed effects and a section-level random intercept. Column (1) reports the estimated standard deviation using the raw estimates. Column (2) shows the correlation of the individual raw PVA estimates from the main and mixed-model specification.

Accuracy of estimated standard deviation of PVA: Estimating the variance of PVA requires separating true heterogeneity from sampling noise. Because the initial estimates of peer effects are noisy, the deconvolution procedure plays a crucial role in recovering the underlying distribution. We evaluate the robustness of our results by altering this deconvolution procedure in several ways.

First, we assess how the computation method that we use to compute standard errors affects our main parameters of interest. Standard errors (SE) are a key input for the empirical Bayes deconvolution process because they directly affect both the estimated prior of PVA and the subsequent shrinkage of individual peer value-added measures. Table A7 reports the estimated SE, the estimated standard deviation (SD) and validation coefficients for a number of different computation methods. The table shows that the clustering choice has a direct impact on our main parameters of interest. Clustering only at the student level, thus ignoring correlated outcomes within a course, leads to larger standard errors, an estimated prior close to zero, and extremely aggressive shrinkage. While it seems questionable to cluster only at the student level, this approach yield forecast-unbiased estimates. In contrast, clustering only at either the course or section level, thus ignoring serial correlation within students, yields SEs that are too conservative. These downward-biased SEs deliver insufficient shrinkage and are not forecast unbiased. Two-way clustering at the student and course level addresses both sources of correlation simultaneously and, by construction, reduces to the correct one-way variance if one source of correlation is in fact absent. While two-way clustering at the student and course level appears the most appropriate from a design perspective, it also delivers the most credible and a forecast-unbiased validation coefficient.

Second, we use a bias-corrected variance estimator following Walters (2024) to recover the variance of PVA without imposing parametric assumptions on the mixing distribution and providing robustness to arbitrary heteroskedasticity. Specifically, we use the efficient approximation for leave-one-out estimates as described in Kline, Saggio and Sølvsten (2020) (KSS) to construct the unbiased variance estimator. The idea of this approach is to learn about the spread of peer effects without explicitly assuming any particular distribution. Table A8 reports the estimated spread of PVA; for raw estimates it assumes our preferred main specification and several adoptions of the KSS approach that differ in assumed error specifications. In Panel A, we estimate the spread of PVA by first regressing course grades on student and course fixed effects and then regress peer fixed effects on the residual. The KSS approach yields results similar to our main empirical Bayes specification when specifying an error structure that is clustered on the student level. Otherwise, the estimated SD cannot be recovered. In Panel B, we jointly estimate peer, student, and course fixed effects.[29] This increases the estimated standard deviation for all specifications. Estimates range between .029 to 0.060 and are larger than our reported main specification.

[29] The one-step approach increases the risk of overfitting the data by estimating all sets of fixed effects jointly. We therefore use the more conservative two-step approach, which might underestimate the true variation as part of the peer influences may be absorbed by student and course fixed effects.

Third, we conduct two nonparametric permutation exercises to (1) provide a nonparametric test of the importance of peers, (2) understand how noisy PVA measures are and (3) test whether cross-validation fails when it should. For exercise one, we randomly reassign students to placebo sections while maintaining the underlying course structure and re-estimate our PVA measures. We then compute the standard deviation of the raw PVA estimates and repeat this permutation exercise 1,000 times. Our approach preserves the structure of our data (e.g., student and course fixed effects and overall variance in outcomes) while breaking any systematic relationship between peer identities and outcomes for the placebo estimates. By comparing our actual estimates to these placebo distributions, we obtain a nonparametric measure of significance that does not rely on parametric assumptions about the underlying data-generating process. Figure A5 show the distribution of the SD of placebo PVA in blue. This distribution highlights substantial variation in PVA estimated in placebo-assigned sections, which is consistent with our low noise-to-signal ratio documented earlier. However, the distribution does not overlap with our true estimate of 0.061 marked with a red vertical line.[30] In exercise two, we apply the same permutation procedure and then use our cross-validation framework to verify whether peer group PVA has predictive power in placebo assignments. Intuitively, this means that we assess whether the estimated peer group PVA can predict course grades in placebo social interactions. Figure A6 shows the distribution of the cross-validation coefficient using either raw or shrunken peer group PVA measures. The cross-validation coefficient distinctively exceeds the placebo distribution in both cases, highlighting that we fail to validate the PVA measures in placebo out-of-sample social interactions. Taken together, our simulation exercises highlight that, while placebo PVA has substantial variation, it has—in contrast to our actual PVA measures—zero predictive power for out-of-sample social interaction.

[30] We report the standard deviation of the raw PVA estimates in the placebo distribution due to computational constraints. Estimating the bias-corrected variance or applying clustered standard errors in each permutation would immensely increase computation time.

Table A7: Comparison Clustering Choices

Error specification	(1) Estimates of PVA SD: Raw estimates	(2) Estimates of PVA SD: Empirical Bayes estimates	(3) Estimates of PVA SD: Latent distribution	(4) Information SE and Noise: Avg. standard error of PVA	(5) Information SE and Noise: Amount original signal	(6) Validation Coefficient: Peer Group PVA
Normal	0.061	0	0	0.067	0	-
Robust	0.061	0	0	0.067	0	-
Cluster student	0.061	0.000	0.005	0.059	0.01	0.943 (0.280)
Cluster section	0.061	0.013	0.027	0.053	0.21	0.468 (0.077)
Cluster course-year	0.061	0.014	0.028	0.052	0.24	0.463 (0.069)
Two-way cluster student and section	0.061	0.018	0.033	0.050	0.31	0.334 (0.060)
Two-way cluster student and course-year	0.061	0.003	0.012	0.058	0.05	1.147 (0.177)
Full bootstrap section level	0.061	0.017	0.031	0.051	0.29	0.281 (0.065)

Note: This table reports key parameters summarizing peer value-added (PVA) across alternative error and clustering specifications. Columns (1)–(3) report the standard deviation of PVA for (1) raw (OLS Peer coefficients), (2) empirical Bayes estimates and (3) the estimated latent distribution. Column (4) reports the mean standard error of the raw PVA coefficients. Column (5) reports the average amount of signal in PVA estimates and is computed by taking the average of $\left(\frac{\hat{\sigma}_Y^2}{\hat{\sigma}_Y^2+s_j^2}\right)$. Column (6) reports the out-of-sample validation coefficient obtained from running a regression of standardized end-of-course grades on Peer Group PVA, which is constructed based on a course-leave-out sample that excludes the course in which we observe the grade we use as a dependent variable. Peer Group PVA sums up the empirical Bayes PVAs of all peers within the same section. We control for the individual empirical Bayes PVA estimate, student-level fixed effects, and course-level fixed effects.

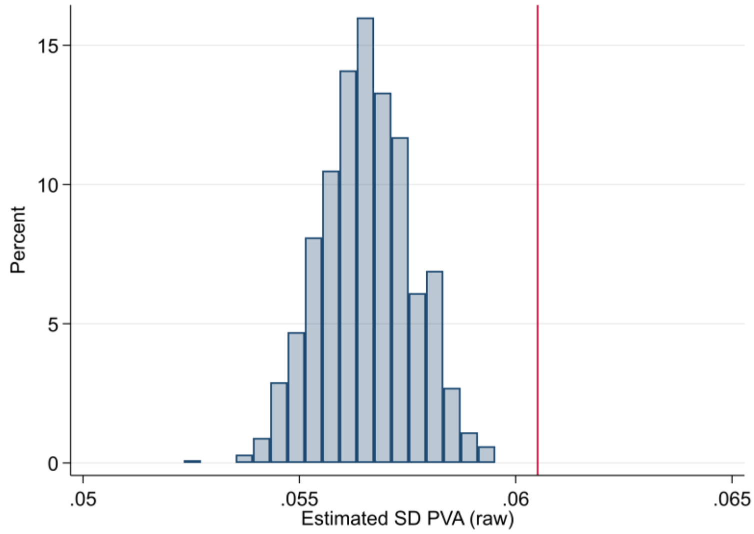
Table A8: Comparison of PVA SD—Empirical Bayes vs. KSS

Panel A: Residualize Student FE and Course FE	(1) SD PVA
Raw estimates	0.061
EB (Two-way clustering student and course level)	0.013
KSS (General)	-
KSS (Clustering student level)	0.011
KSS (Clustering course level)	-
KSS (Clustering section level)	-

Panel B: Full Model Student FE and Course FE	SD PVA
Raw estimates	0.102
EB (Two-way clustering student and course level)	0.067
KSS (General)	0.040
KSS (Clustering student level)	0.029
KSS (Clustering course level)	0.049
KSS (Clustering section level)	0.060

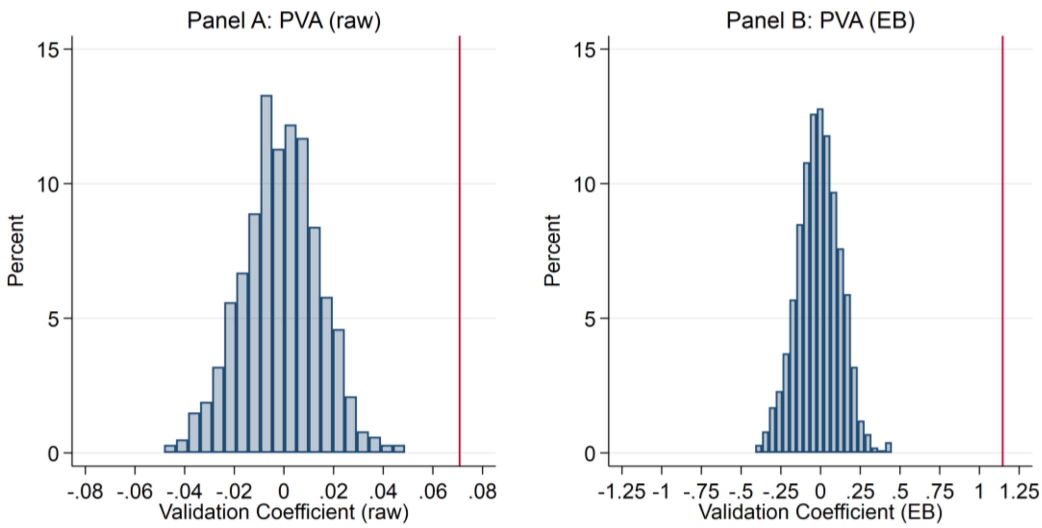
Note: This table reports the standard deviation (SD) of peer value-added (PVA) under empirical Bayes and Kline, Saggio and Sølvsten (2020) (KSS) estimators. Panel A computes PVA by first residualizing standardized end-of-course grades using student-level and course-level fixed effects. The residualized end-of-course grades are then the input for estimating individual PVA and standard deviation using empirical Bayes or KSS estimators. Panel B estimates the full model with student-level and course-level fixed effects included directly. KSS (General) uses the leave-one-out variance-component estimator with a heteroskedastic-robust error structure. For KSS (Clustering) estimators allow arbitrary covariance within the defined cluster.

Figure A5: Placebo Peer Assignments—Distribution of PVA



Note: This figure shows the distribution of estimated SDs of raw PVA using placebo reassignments. We simulate 1,000 random reassignments of students within courses into the existing sections and estimate PVA using the placebo reassignments. The red line indicates the standard deviation of PVA estimates using the true assignment.

Figure A6: Placebo Test—Predictive Power of PVA

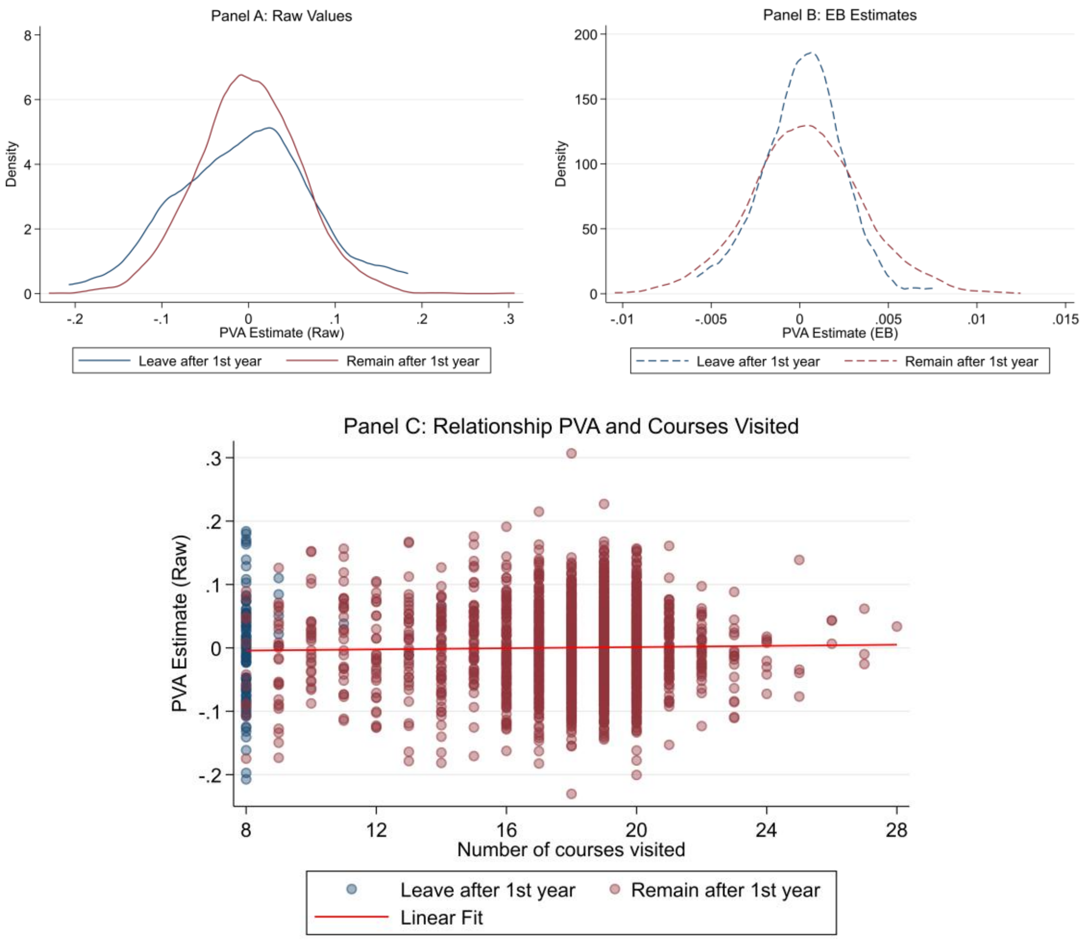


Note: The figure shows the distribution of the validation coefficient using Peer Group PVA estimates in 1,000 randomly reassigned students within courses. Peer Group PVA is constructed based on a course-leave-out sample that excludes the course in which we observe the grade we use as a dependent variable. Peer Group PVA sums up the estimated PVAs of all peers within the same (reassigned) section. We control for the individual PVA estimate, student-level fixed effects, and course-level fixed effects. Panel A shows the distribution of validation coefficients when using raw PVA estimates, and Panel B reports the distribution of validation coefficients when using empirical Bayes PVA estimates. The red line indicates the validation coefficient obtained by using the true assignment.

Attrition: Our estimation relies on an imbalanced sample of student-section observations, as not all students are observed for the entire three-year program. We observe 63.9 percent of students for at least the full three years. We observe 28.8 percent of students for only one year, most of whom are exchange students or enrolled in other study programs. When conditioning on students who were at some point part of the two main study programs, 15.6 percent are observed in only one year. Better-performing students are observed in more courses.

How does this type of attrition affect the peer value-added distribution that we observe? One concern could be that with fewer observations per peer, raw fixed-effect estimates become noisier, widening the empirical distribution. It would also be important to know if early leavers happen to have extreme (true) PVA, implying that dropping them could thin the tails of the latent distribution. We therefore investigate whether PVA differs for students classified as "leavers" and "stayers." Students are classified as leavers if they only appear in first year courses. All others are classified as stayers. Figure A7 shows the distribution of raw and shrunken PVA in Panels A and B. Perhaps unsurprisingly, raw estimates are more dispersed for leavers, likely due to being estimated with a lot fewer observations per student. This pattern reverses for the shrunken values, highlighting that the individual estimates of leavers are inherently noisier. The difference between average PVA of stayers and leavers is not significant. Panel C further shows that estimated raw PVA does not systematically differ by the number of course observations per student. Taken together, these results highlight that attrition does not seem to introduce a systematic bias into the PVA distribution and that it does not seem to create meaningful distortions in the observed distribution of PVA.

Figure A7: PVA of Leavers and Stayers



Note: These figures show the distribution of PVA split by students that are classified as leavers and stayers. Students are classified as leavers if they only appear in first-year courses. All others are classified as stayers. Panel A shows the distribution of the raw PVA estimates. Panel B shows the distribution of the empirical Bayes PVA estimates. Panel C shows the relationship between the estimated raw PVA and number of course observations per student. The red line indicates the linear fit between individual raw PVA and number of course observations per student.

A2.2 Performance of PVA in Monte Carlo Simulations

Simulation evidence: To evaluate under which conditions our empirical strategy can recover peer value-added (PVA) variance, we conduct a series of Monte Carlo simulations, addressing two critical questions. First, can our sample size and data structure support credible identification of PVA variance despite the limited rounds of reassignment? Second, can mixed-effects models accurately separate peer and section-level variance components, particularly in the presence of unobserved classroom-level shocks?

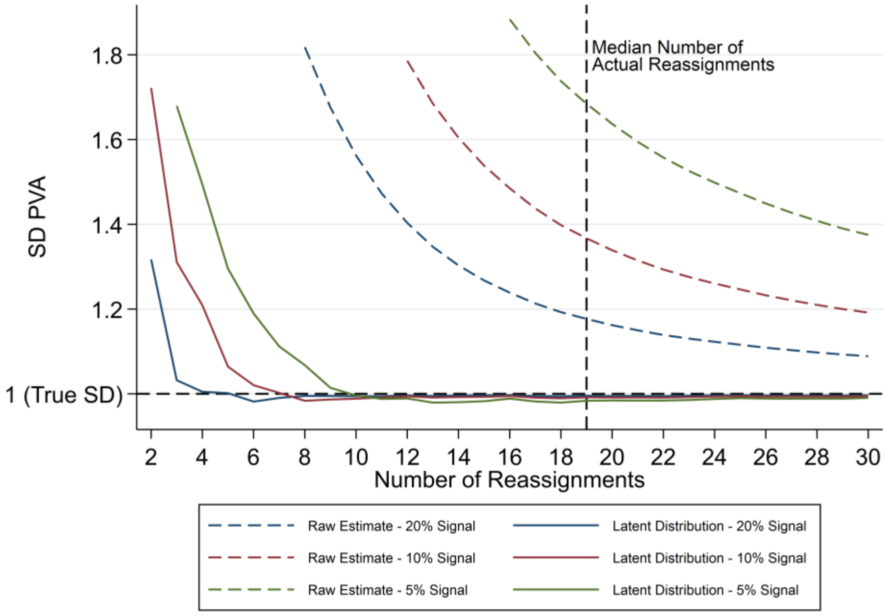
For the first question, we design a simulation with 600 students that are repeatedly assigned to peer groups of various peer group sizes across multiple courses. Performance outcomes are generated as a combination of peer effects and i.i.d. noise, with varying variation in the signal-to-noise ratio. We estimate peer value-added using both raw fixed-effects and empirical Bayes methods, and track the recovery of the true standard deviation of the latent PVA distribution as a function of the number of reassignments and section size. Panel A of Figure A8 shows how the raw variance and estimated hyperparameter converge to the true variance of PVA with increasing number of reassignments, holding group size constant at 12 (which is the median group size in our setting). Figure A8 shows that the empirical Bayes hyperparameter quickly converges under the assumed DGPs. With moderate section sizes and more than ten reassignments per peer we can reliably recover the true variance of PVA. Convergence is notably slower for the variance of raw PVA estimates, highlighting that the size of raw estimates should be interpreted with caution under a limited number of reassignments, highlighting the need for empirical Bayes shrinkage.

Figure A9 shows convergence of PVA as section size increases, holding the number of reassignments constant at 19 (our observed median number of peer reassignments). As section size increases, the raw variance estimates of PVA become increasingly inflated, with the degree of inflation depending on the signal strength. In contrast, the empirical Bayes estimates remain close to the true PVA variance even with large section sizes, demonstrating that the empirical Bayes hyperparameter estimate converges reliably to the true PVA variance even with large section sizes.

For the second question regarding estimating PVA with mixed models, we maintain our actual empirical data structure and simulate student-level PVA, i.i.d.

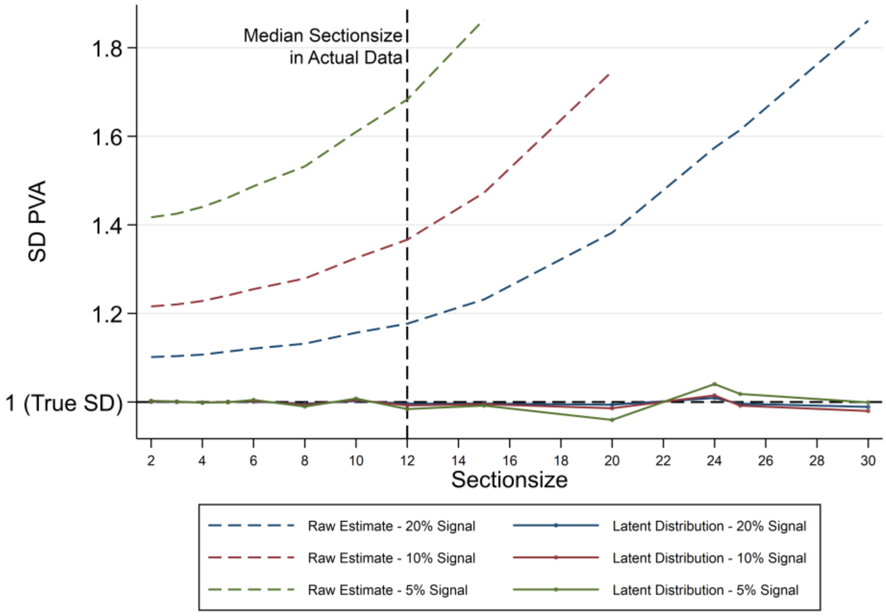
noise, and section-level shocks. We keep the true PVA variance constant while varying the relative contribution of i.i.d. noise and section-level noise across outcomes. This allows us to examine how OLS and mixed-effects model estimates respond as section-level noise becomes more important, and whether mixed-effects models can correctly attribute variance to the peer level rather than to the section level. Our results, summarized in Figure A10, show that when no section-level shocks are present, both OLS and mixed-effects models recover the true standard deviation of PVA accurately and yield nearly identical estimates. As the importance of section-level shocks increases, mixed-effects models continue to recover the true peer variance reliably, while the OLS estimates become increasingly inflated. These results indicate that the estimated variance in our empirical data is not driven by unobserved section-level shocks.

Figure A8: Impact of Assignment Frequency on Recovering Variation in PVA



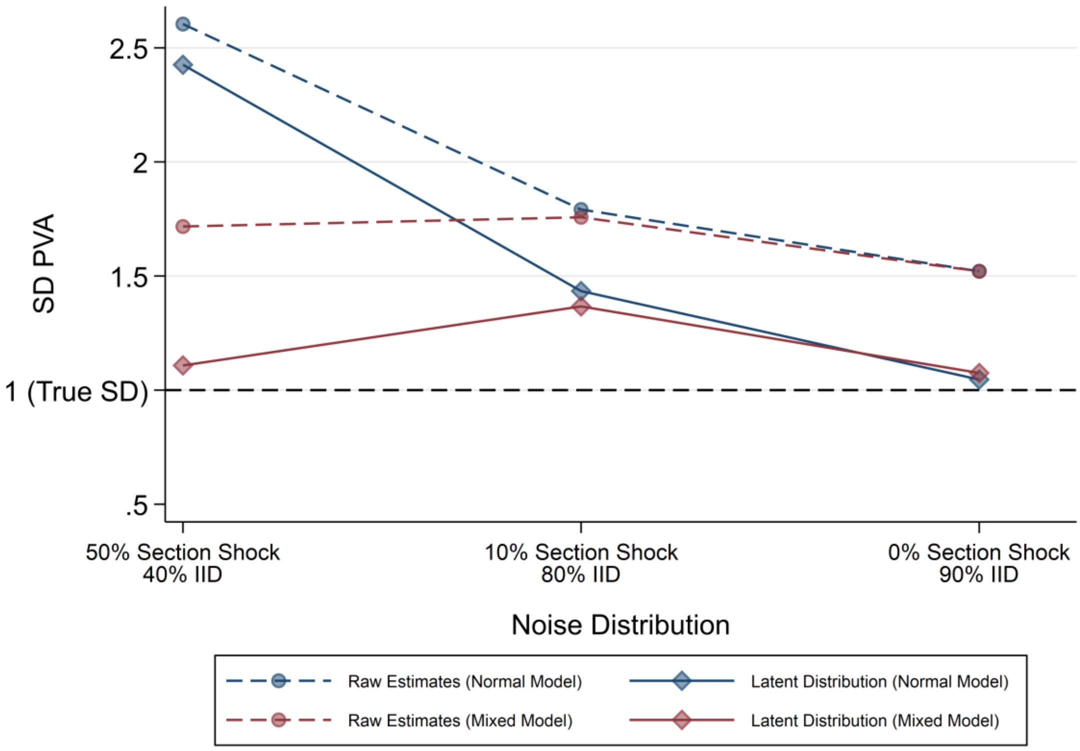
Note: This figure shows the convergence of estimated SD of PVA for increasing the number of reassignments of peers into sections using Monte Carlo simulation. The simulation features 600 students and constant section sizes of 12 students per section. For each course, students are randomly assigned to a corresponding section. Estimated PVA is computed based on simulated performance outcomes Y^1, Y^2, Y^3, where $Y_{ics}^k(G_{ics}) = \Theta_{ics} + \sqrt{(1-\rho^k) * (Var(\Theta_{ics})/\rho^k)} * \epsilon_{ics}$ and $\rho^1 = 0.2, \rho^2 = 0.1, \rho^3 = 0.05$, that is. the outcomes contain 20%, 10% and 5% PVA signal. The solid lines indicate the average estimated SD when computing the SD of the estimated latent distribution. The dashed lines indicate the average SD computed by using the raw estimates. The straight dashed black line indicates the median number of reassignments for students we use to estimate PVA in our actual dataset. Monte Carlo results are based on 100 repetitions. Values bigger than 1.9 are not displayed.

Figure A9: Impact of Section Size on Recovering Variation in PVA



Note: This figure shows the change of estimated SD of PVA when increasing the section size of peers into sections using Monte Carlo simulation. The simulation features 600 students and 19 courses. We use constant section sizes of 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 24, 25, and 30. For each course, students are randomly assigned to a corresponding section. Estimated PVA is computed based on simulated performance outcomes Y^1, Y^2, Y^3, where $Y^k_{ics}(G_{ics}) = \Theta_{ics} + \sqrt{(1-\rho^k) * (Var(\Theta_{ics})/\rho^k)} * \epsilon_{ics}$ and $\rho^1 = 0.2$, $\rho^2 = 0.1$, $\rho^3 = 0.05$, that is. the outcomes contain 20%, 10% and 5% PVA signal. The solid lines indicate the average estimated SD when computing the SD of the estimated latent distribution. The dashed lines indicate the average SD computed by using the raw estimates. The straight dashed black line indicates the median section size in our actual dataset. Monte Carlo results are based on 100 repetitions. Values bigger than 1.9 are not displayed.

Figure A10: Comparing Mixed-Model Approach



Note: This figure shows the change of estimated SD of PVA when changing the relative importance of i.i.d. to section-level noise using Monte Carlo simulation. The simulation features the *original* dataset and uses simulated values for PVA and i.i.d. noise and section noise. Individual PVA, i.i.d noise, and section noise are all drawn from independent normal distributions N(0, 1). Estimated PVA is computed based on simulated performance outcomes, in which we keep PVA signal constant at 10% and vary the signal strength of the i.i.d. noise and section noise. For each outcome we estimate PVA using either OLS (only peer dummies) or a mixed model in which we additionally specify to model the section variance component. The solid lines indicate the average estimated SD of the latent distribution using empirical Bayes. The dashed lines indicate the average SD computed by using the raw estimates. Monte Carlo results are based on 25 repetitions.

A2.3 Match Effects

Match effects arise when the influence of a peer depends on the characteristics of the student they interact with, such as shared nationality, gender, or ability. These effects violate the constant-effects assumption of our framework, which posits that each peer exerts the same influence on all students regardless of context.

We distinguish three base-case scenarios of how match effects may shape the distribution of PVA:

(i) Distinct PVA, same distribution: Each student has two distinct and uncorrelated PVA values, one for same-group and one for different-group peers, both drawn from the same distribution.

(ii) Mean shift in distributions: The spread of PVA is identical across groups, but the mean is higher when peers share the characteristic.

(iii) Different distributions: Students exert less influence when assigned to peers without the shared characteristics, for example, due to less interaction or ignorance.

The top row of Figure A11 visualizes these base-case scenarios with simulated PVA distributions. Panel A depicts a situation in which each peer exerts two distinct value-added effects—one for same-group matches and another for different-group matches—both drawn from the same distribution. Accordingly, distributions of PVA toward same and different group members strongly overlap. Panel B depicts a situation in which match effects involve a mean shift: peers are systematically more (or less) influential when matched with students sharing certain characteristics, resulting in two distributions identical in shape, but horizontally shifted. Panel C depicts a third situation in which peers interact less effectively across group boundaries. This case results in different distributions of value-added, with a less dispersed distribution for different- than same-group interactions.

If alignment of student and peer characteristics matter in our context, individual PVA estimates will effectively reflect each peer's average influence across all matches, rather than their potentially varying effects on different students. This has direct implications for the estimated variance of peer effects when using a constant-effects framework. In the simple binary case with match-specific effects, the average PVA is determined by

$$\overline{\gamma_J} = \pi \cdot \gamma_j^{\text{same}} + (1 - \pi) \cdot \gamma_j^{\text{diff}}, \quad (15)$$

where π and $(1 - \pi)$ denote the share of same-group and different-group matches. The variance of average PVA is given by

$$\text{Var}(\overline{\gamma_J}) = \pi^2 \sigma_{same}^2 + (1 - \pi)^2 \sigma_{diff}^2 + 2\pi(1 - \pi)\rho\sigma_{same}\sigma_{diff}. \quad (16)$$

Here, σ_{same}^2 and σ_{diff}^2 are the true variances of same-group and different-group PVA, and ρ is the correlation coefficient between same- and different-group PVA. From this we can derive several implications that arise when using the constant-effects framework.

First, the estimated variance from the constant-effects framework is a lower bound on the variance of the true match-specific peer effects, unless all same-group and different-group effects are perfectly correlated. When match effects are uncorrelated or negatively correlated, the averaging leads to attenuation of the recovered variance. Second, the magnitude of the attenuation depends on both the correlation ρ and the weights π. The closer ρ to 1, that is. match effects are perfectly aligned, the smaller the attenuation. When $\rho = -1$ and $\pi = 0.5$, the estimated variance can collapse to zero, meaning PVA appears homogenous despite strong underlying heterogeneity. Third, the estimated variance under the constant-effects framework equals the variance of the true variance if and only if $\rho = 1$ and $\sigma_{same}^2 = \sigma_{diff}^2$. This implies a pure mean shift scenario with no change in dispersion or shape. In this case, the average preserves all variation. These results highlight that the constant-effects framework captures only average peer contributions and may lead to understating the true variance in peer effects.

An alternative specification would allow for PVA to differ by match characteristics by adjusting equation (6) to

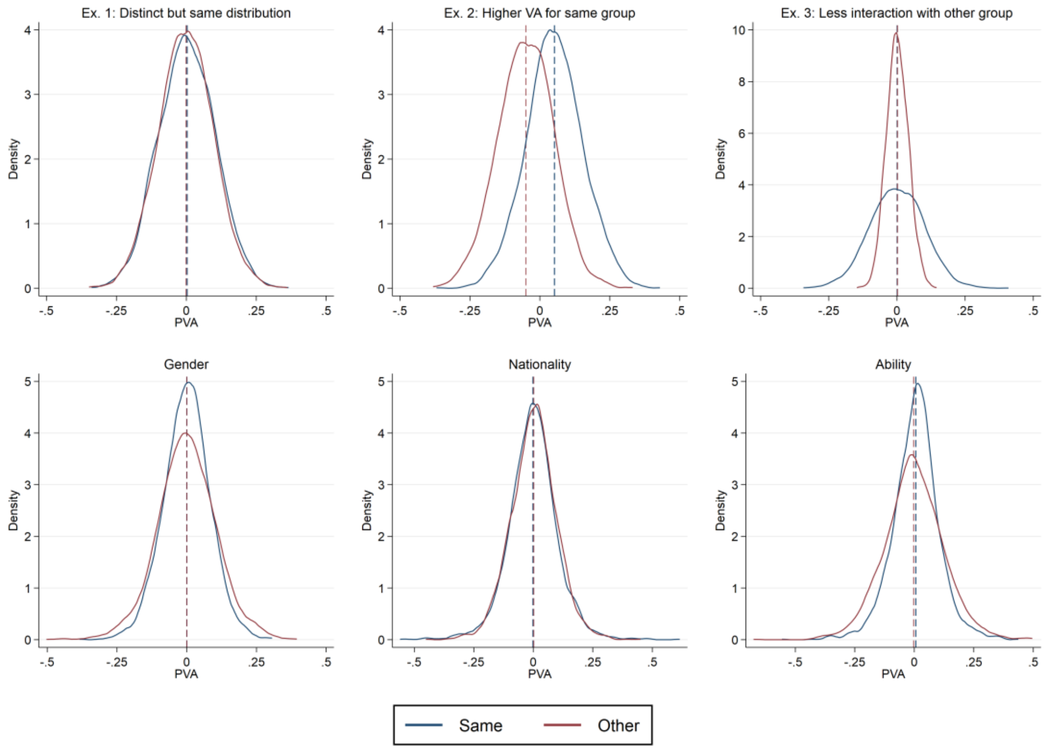
$$\hat{u}_{ics} = \beta_0 + \sum_{j \neq i}(\gamma_j^{same} \times D_{ijcs} \times M_{ij}^g + \gamma_j^{diff} \times D_{ijcs} \times (1 - M_{ij}^g)) + \varepsilon_{ics}, \quad (17)$$

where M_{ij}^g is a binary variable indicating whether a peer j and a student i belong to the same group g (i.e., same gender, ability, or nationality). γ_j^{same} captures peer j's effect when matched to a student from the same group based on characteristic g, and γ_j^{diff} capture peer j's effect when matched to a student from the different group. Note that this specification is considerably more demanding, effectively doubling the number of parameters to be estimated, and substantially reducing the number of observations available for the identification of each parameter. In addition, this requires pre-specification of the relevant matching dimensions and reliance on observable characteristics to define them, alas limiting and potentially missing match heterogeneity driven by unobserved or misclassified traits.

We first provide simulation exercises as proof of the concept of how ignoring different forms of match effects impact our PVA variance estimate. We do so by extending our Monte Carlo simulation from Section 4.4 and assign each student a randomly assigned binary characteristic. We use these characteristics to assign same- and different-group peer status and generate different forms of PVA following the three base-case scenarios discussed above. Table A9 summarizes the SD estimates of PVA when using either the constant-effects or match-specific framework. In all three match-effect scenarios, and in line with the theoretical predictions, the constant-effects approach recovers substantially lower estimates of PVA variance than the true value (set to one in the DPG). In contrast, the match-specific framework successfully recovers the full variance of peer effects.

Building on this simulation, we next apply the match-specific framework to our actual data, focusing on three commonly studied dimensions of matching: gender, ability, and nationality. Results, shown in the lower panels of Figure A11, do not suggest strong evidence for match effects. We do *not* observe a mean shift between PVA estimates for same-group students and different-group students for any of the three dimensions. For gender and ability, PVA appears to be slightly more dispersed for the same than for the different group. For nationality, distributions are strongly overlapping.

Figure A11: Type of Match Effects



Note: This figure visualizes match-specific peer value-added (PVA) distributions in which we distinguish PVA based on a shared or different student-level characteristic. The top row shows examples of different match-specific PVA frameworks. The bottom row shows the estimated raw PVA values when specifying a regression in which we define separate peer dummies for same and other shared characteristics along the dimensions of gender, nationality, and ability. For ability, we split the sample in low and high ability based on a median split of the end-of-course grades. The dotted lines show the respective mean value of the distributions.

Residual sensitivity check: One of the most prominent dimensions of match effects in the literature on peer effects is a students' own ability. Canonical peer effects studies often estimate peer effect specifications that allow for heterogeneity by own and recipients' ability (Carrell, Sacerdote and West (2013), Carrell, Fullerton and West (2009), Burke and Sass (2013)), often finding nonlinear effects that go beyond the simple same/different group distinction examined before. We thus corroborate our analysis with a second exercise following Card, Heining and Kline (2013) that examines residuals of equation (6) by deciles of students' own ability (defined by student fixed effects) and deciles of PVA. Homogeneity in residuals across these cells supports the identification assumptions of separative additivity and constant effects. Figure A12 shows average residuals by deciles of peer and student quality. For the bulk of the sample, deviations remain small, below 5 percent of a standard deviation. Largest deviations are found for very low-ability students who apparently are particularly harmed by being exposed to low value-added peers but also gain more from being exposed to high value-added peers. Vice versa, very high ability students are harmed less by being exposed to low value-added peers, but also gain less from being exposed to high value-added peers.

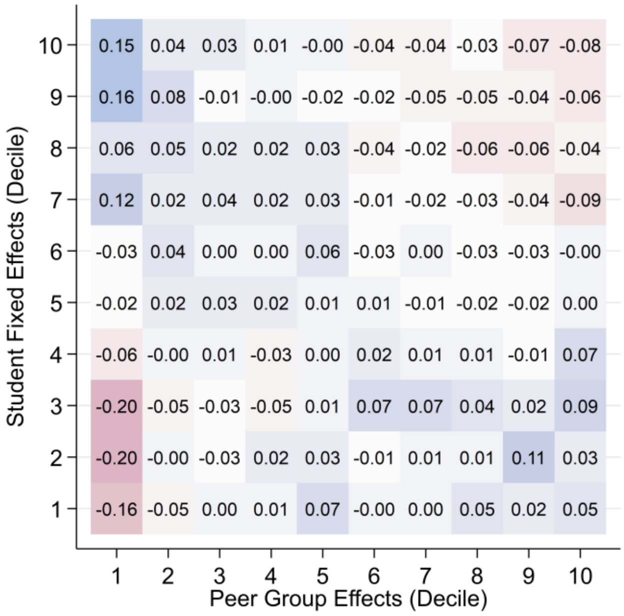
While our simulations show how match effects can bias estimates of PVA, our empirical evidence does not reveal clear or systematic patterns of match-specific heterogeneity. Nevertheless, the limited number of reassignments and the large number of parameters make us underpowered to detect more subtle forms of match effects. These limitations should be kept in mind when interpreting our variance estimates.

Table A9: Match Effects Simulation

		Constant-effects framework	Match-specific Effect Framework	
Classification	DGP	Estimate SD (Overall)	estimate SD (same group)	estimate SD (different group)
No match effects	PVA = N(0, 1)	0.99	0.96	1.01
Distinct PVA, same distribution	PVA (Same) = N(0, 1) PVA (Different) = N(0, 1)	0.70	0.98	1.00
Higher VA for same group peers	PVA (Same) = N(1, 1) PVA (Different) = N(0, 1)	0.71	0.98	1.00
Different distributions	PVA (Same) = N(0, 1) PVA (Different) = N(0, .5)	0.55	0.98	0.51

Notes: This table presents simulation evidence on how different types of match effects affect the estimated standard deviation (SD) of peer value-added (PVA). The simulation features 600 students, constant section sizes of 12 students per section and 20 courses. Performance outcomes are generated using a prespecified data-generating processes with and without match effects. We estimate PVA using both a constant-effects framework (column 3) and a match-specific framework with separate peer indicators for same-group and different-group interactions (columns 4–5). All reported SD estimates refer to the empirical Bayes hyperparameter estimate of the latent distribution. Monte Carlo results are based on 100 repetitions.

Figure A12: Residual Sensitivity Check



Note: Following Card, Heining and Kline (2013), this figure shows average residuals of a regression of student grades on course- student- and peer-fixed effects by deciles of the student- and peer group effects distribution. The values indicate the mean residuals per student and peer group cell.

APPENDIX A3: *Peer Value-Added and Linear-In-Means Models*

In this section, we describe the relationship between the PVA approach and canonical linear-in-means (LIM) models of peer effects. In its simplest form, LIM models of peer effects relate an individual outcome y_i to a leave-out mean of one observed peer characteristic x like peer GPA, peer gender, or peer race. Consider the LIM model in equation (18)

$$y_i = \beta \bar{x}_{-i} + \epsilon_i, \tag{18}$$

where β captures the effect of changing $\bar{x}_{-i}$, the peer group average of x excluding i, on outcome y_i. This model captures effects that are driven by x; effects that are driven by other peer characteristics are absorbed by the error term ϵ_i. In contrast, PVA estimates the individual contribution of each peer to others' outcomes, without restricting peer influence to operate only through x.

Below is a simplified version of our potential outcome equation,

$$Y_{ics}(G_{ics}) = \sum_{j \in G_{ics}} PVA_j + \epsilon_{ics} \tag{19}$$

in which we define potential Y_{ics} as the sum of the peer value-added of all peers in section s and a random error ϵ_{ics}. We assume that each peer's PVA is determined by the observed characteristic x and an unobserved characteristic z:

$$\mathrm{PVA_j} = a\, x_j + (1 - a)\, z_j,$$

with

$$\mathrm{Var}(x) = \sigma_x^2, \quad \mathrm{Var}(z) = \sigma_z^2, \quad \mathrm{Cov}(x, z) = 0.$$

The LIM coefficient can then be expressed as

$$\beta = \frac{\mathrm{Cov}(\overline{x}_{-i}, y_i)}{Var(\overline{x}_{-i})} = \frac{a\, \sigma_x^2}{\sigma_x^2 / G} = a\, G,$$

where G is the number of peers in peer group G. We can further express a as

$$a = \frac{\text{Cov}(x,\text{PVA})}{\text{Var}(x)} = \frac{\text{Cov}(x,\text{PVA})}{\sigma_x^2}.$$

substituting into β gives

$$\beta = G \cdot \frac{\text{Cov}(x,\text{PVA})}{\text{Var}(x)}.$$

This equation reveals a key insight: the LIM estimate captures the correlation between PVA and the observed peer characteristic x. If PVA is uncorrelated with x, LIM estimates do not detect a peer effect—even when meaningful peer effects exist.[31] In our specific setting, none of our observed student characteristics predict PVA (see Table 4). This implies that none of the conventional LIM estimates would pick up any significant effects. However, our PVA estimates reveal meaningful variation in individual peers' contributions to performance. These results highlight how LIM estimates can miss meaningful peer effects.

[31] On the flip side, if a peers' x does not affect outcome, but it is correlated with PVA, the LIM estimator will detect peer effects. For example, positive LIM estimates of the effect of having higher-GPA peers might not be driven by peer GPA but by the correlation of peer GPA with PVA.