The Value of a Peer[*]

Ingo E. Isphording
(IZA)

Ulf Zölitz
(University of Zurich)

June, 2022

Abstract

This paper introduces peer value-added, a new approach to quantify the total contribution of an individual peer to the performance of others. We estimate peer value-added using student- and peer- two-way fixed effects models. Using data with repeated random assignment of students to university sections we find that peers differ significantly in their value-added: A one standard deviation better peer increases performance by 0.037 standard deviations. To validate our measure of peer quality, we show that peer value-added predicts student performance out of sample. We find that observable peer characteristics, including peer achievement, are poor predictors of peer value-added. Our findings highlight that the canonical approach to estimating peer effects misses spillovers unrelated to observable peer characteristics.

Keywords: peer effects, peer value-added, peer capital, spillovers

JEL classification: I21, I24, J24

[*] An earlier version of this paper was circulated as University of Zurich Working Paper No. 342, 2019. We thank our valuable peers Steffen Altmann, David Dorn, Jan Feld, John N. Friedman, Daniel Hamermesh, Jonas Radbruch, Nicolás Salamanca, 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. Ingo E. Isphording: IZA – Institute of Labor Economics, Schaumburg-Lippe-Str. 5-9, 53113 Bonn. Ulf Zölitz: University of Zurich, Department of Economics and Jacobs Center for Productive Youth and Child Development, CESifo, CEPR, IZA, and Maastricht University, Schönberggasse 1, 8001 Zurich. ulf.zoelitz@econ.uzh.ch

1. Introduction

Peers matter. In the classroom and the workplace, we learn from each other and become more productive if we meet the right peers. But who are the "right" peers, and how do we quantify who brings out the best in us? The quest to identify valuable peers has motivated a large body of literature studying which peer characteristics affect student performance. However, after decades of research, the literature has not yet reached a consensus on the size of peer effects, which peer characteristics matter, or how to estimate peer effects (Sacerdote 2014).

In this paper, we introduce a new concept to quantify the importance of peers. The *peer value-added* of an individual peer captures their contribution to the performance of others. Peer value-added summarizes *any* spillover on others' performances, regardless of whether spillovers stem from observable or unobservable peer characteristics. Peer value-added can be estimated even when information on observable peer characteristics is absent. While peer value-added is a person-specific measure, it can be aggregated to *peer capital*: Peer capital measures the value-added of all peers an individual student has been exposed to.

We estimate peer value-added in a business school where students are repeatedly and randomly assigned to sections of 10 to 16 peers. Based on 916,842 dyadic student-peer interactions we estimate a two-way student- and peer-fixed effects model. Peer-fixed effects capture the average performance change of students who met a specific peer and measure peer value-added. We shrink these peer value-added measures towards the sample mean to increase their out-of-sample predictive power when used as an explanatory variable.

We document five sets of results. First, peer value-added – the ability to improve fellow students' performance – significantly differs between peers. Being randomly assigned to a peer with a one standard deviation higher value-added increases a student's grade by 3.7 percent of a standard deviation. Most peers only have small impacts on performance – only 16 percent of all peers exert an influence of more than 5 percent of a standard deviation on grades.

Second, we show that peer value-added contains meaningful information on peer-specific spillovers. In our cross-validation exercise, we show that a leave-one-out jackknife measure of peer value-added predicts student performance in interactions that were not part of the initial estimation sample. When we regress performance on our jackknife measure of peer value-added we obtain a significant coefficient close to one. This validation is a prerequisite for deriving valid policy recommendations from peer effects estimates. Carrell, Sacerdote and West (2013) show that standard peer effects estimates can be invalid out-of-sample predictors and lead to suboptimal group assignments.

Third, we quantify the overall importance of a pool of peers. We construct *peer capital* as the average (jackknife) peer value-added of all peers that a student meets throughout their studies. We then relate a student's peer capital to their performance. We find that "lucky" students from the highest quartile of the peer capital distribution have a 6 percent of a standard deviation higher grade point average (GPA) than "unlucky" students with a peer group from the lowest quartile of the peer capital distribution.

Fourth, we study the impact of special peers at the tails of the peer value-added distribution: We define *shining lights* and *bad apples* as peers whose value-added falls into the highest and lowest deciles of the peer value-added distribution. We then test how the number of bad apples and shining lights affects performance. We find that the number of these special peers has an approximately linear effect on performance – with one exception: Having only one "shining light" in a section does not create any positive spillovers. This finding is consistent with exceptionally good peers requiring a sparring partner in class to generate positive spillovers on other students.

Fifth, we provide insights into the correlates and stability of peer value-added. We show that observable peer characteristics, including peer achievement, are poor predictors of peer value-added. This result runs contrary to the popular belief that higher-achieving students produce the largest spillovers. To understand how universal peer value-added is, we construct value-added measures by gender, ability, and nationality subgroups and document substantial overlap in how a specific peer affects performance across these groups. These results suggest that peer value-added has a common core, which produces spillovers in different types of social interactions.

With this paper, we make conceptual and empirical contributions 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 relates a group-level aggregate of a single observable characteristic to student outcomes. This literature studies, for example, how student performance is affected by having higher-achieving peers, more black peers, more female peers, or more free-lunch peers.[1] In contrast to our framework, these studies are motivated by the

[1] 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 more

idea that the ability to raise others' performance is closely related to one's observable characteristics (Fruehwirth 2014). By recognizing that spillovers can arise from both observed and unobserved peer characteristics, we propose a more comprehensive approach to estimating peer effects. Our approach does not rely 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, however, 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 person-specific peer value-added. The poor correlation between peer value-added and observable peer characteristics indicates that social spillovers are not well captured by student observables. The canonical approach of estimating peer effects misses spillovers that are weakly correlated with observable peer characteristics.

Our approach to identify peer value-added is inspired by the literature on teacher value-added.[2] 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 is three to five times larger than the variation in peer value-added. This shows, perhaps unsurprisingly, that differences between individual teachers affect performance more than differences between individual peers. Despite the relatively small effect of a single peer, we demonstrate that the entirety of peers has substantial aggregate effects on student performance.

The remainder of the paper is structured as follows. In Section 2, we lay out the conceptual framework of peer value-added. In Section 3, we describe the institutional details and data we use. Section 4 describes how we estimate peer value-added. In Section 5, we present results on variation, predictive power, and correlates of peer value-added. Section 6 discusses how our findings connect to the existing literature and potential applications. Section 7 concludes.

---

recent evidence on ability peer effects.[2] Koedel et al. (2015) and Hanushek and Rivkin (2006) provide a comprehensive review on the extensive literature on teacher value-added.

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

2. Conceptual Framework

We aim to quantify the total contribution of an individual peer to the performance of their fellow students. We define *peer value-added* (PVA) as the average effect of an individual peer on the performance of their fellow students.

The PVA approach differs from the existing literature on peer effects in two important ways. First, PVA identifies the *person-specific* contribution to performance. It describes the expected change in student performance caused by being in class with one specific peer relative to meeting a peer with average value-added. In contrast, the standard peer effects literature focuses on the contribution of an entire *group of peers*, where group contributions are estimated using observable peer characteristics. Frequently used characteristics include average pre-assignment test scores, the share of female peers, the share of black peers, or the share of free-lunch-receiving peers. This aggregation to the group-level makes it impossible to isolate the contribution of an individual peer. Second, PVA measures the *total contribution* to others' performance. The total contribution includes spillovers due to both observable and unobservable peer characteristics. With the PVA approach, there is no need to form ex-ante hypotheses about which observable dimensions generate spillovers. PVA can be estimated even when data on student characteristics are absent. This distinction is particularly important when peer effects are driven by unobserved differences between students. In contrast to our approach, standard peer effects studies can only identify spillovers in dimensions that are sufficiently correlated with observable characteristics. These studies might therefore falsely conclude that peer effects are small or non-existent when observable characteristics are not sufficiently correlated with spillovers.

We derive the *person-specific total contribution* from a simple education production framework that includes the peer environment as a separate input factor. Borrowing notation from Angrist et al. (2017), we consider a population of N students repeatedly interacting with each other in S peer groups (sections). Spillovers in educational production arise from *interactions* between student i and peer j. The potential outcome Y_{ijs} is the performance of student i when meeting peer j in section s. It is described by a constant-effects-model with three non-interacting components:

$$Y_{ijs} = \alpha_i + \gamma_j + \mu_s \qquad (1)$$

where α_i is a student i's time-invariant latent ability, γ_j is the contribution of peer j, μ_s captures factors that can be attributed to the section. The observed outcome of student i who meets peer j in section s can be written as

$$Y_{ijs} = \gamma_0 + \sum_{j=1}^{J} \beta_j D_j + \alpha_i + \mu_s \tag{2}$$

where D_j is an indicator for the presence of peer j in section s. J is the total number of peers. Our parameter of interest, β_j, captures peer *j's* value-added as $\gamma_j - \gamma_0$, i.e., the causal effect of peer *j's* presence relative to a peer with average value-added.

Variation in peer value-added can arise from differences in peer behaviors, skills, abilities, or traits. For example, pro-social peers might deliberately support other students, while other peers might have skill sets that are complementary to students' learning progress. Peers might also act as role models and inspire their fellow students to work harder. Furthermore, some peers might possess specific personality traits that generate a productive learning environment for their fellow students. These mechanisms can also be framed negatively: students can be disruptive, possess a non-complementary skill set, discourage other students from working hard, and generate a hostile learning environment through their personality. Irrespective of the exact mechanism, peer value-added, β_j, captures the expected contribution of peer j to the performance of student i.

The constant effects model in equations (1) and (2) implies that each student in a section is equally affected by the presence of peer j. However, the ability to raise others' performance might also depend on the specific peer composition or class content. We later explore heterogeneity across different subjects, gender, ability, and nationality.

3. Institutional Environment, Data, and Randomization

We estimate PVA with data from a Dutch business school where 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 assignment process as a key feature of our empirical strategy.[3]

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

3.1 Institutional Environment

We use data from a Dutch business school that offers bachelor's, master's, and Ph.D. programs. Our analysis focuses on two large bachelor programs in which all first-year students follow the same general course structure and take the same 8 compulsory courses. In addition, a typical student takes 10-12 compulsory and elective courses in the 2nd and 3rd years and spends one term abroad.

We have access to data from six academic years between 2009–10 and 2014–15. Teaching at the business school takes place in four regular periods per academic year. Each of those teaching period lasts about two months. In first-year compulsory courses, students sit centrally graded written exams at the end of each period and this exam grade is the only assessment. In addition to weekly lectures, students meet in randomly assigned sections of 10 to 16 students. For each course, students are randomly assigned to a new 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 meeting with their section peers to discuss the material. These discussions are monitored by an instructor, who can be a professor, lecturer, graduate, or undergraduate student (Feld et al. 2018).

A key feature of the business school is that the scheduling office assigns students and instructors within courses to sections on a random basis. The random assignment is a by-product of the scheduling software used at the institution. Neither students nor instructors interfere with this process. Since 2011, the random assignment has been stratified by nationality.[4] We exclude sections from the analysis where course coordinators did not comply with the standard allocation procedure. In Section 3.3 we present balancing checks confirming the random assignment of students to sections.

The assignment of students to sections is binding. Switching from the assigned section is only allowed for medical reasons, or when the student is a top athlete and must attend sports practice. Students are required to attend their designated section. To be admitted to the exam, they

[4] 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 included 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.

must not miss more than three meetings of their designated section. Instructors keep a record of attendance. The attendance data are not centrally stored and thus not available to us.

3.2 Data

To estimate the peer-specific contribution to students' performance (PVA) we rely on *dyadic* student interaction data: The data follows a student-peer-course-section-level ($i \times j \times c \times s$) data structure, in which each *interaction* between a student i and a peer j in section s in course c constitutes a single observation. Students cannot interact with themselves, so $i \neq j$ in all interactions. This dyadic data allows us to estimate independent contributions of single peers, instead of focusing on group effects.

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, we compute pre-assignment student GPA based on all grades received before a section assignment takes place. This pre-determined student GPA is necessarily missing for the first teaching period.

Descriptive statistics: Table 1 describes the dataset. We observe 9,825 students in six cohorts. 39 percent are female, and students are on average 20 years old. 23 percent of students are Dutch and 38 percent are German. On average, students score 6.57 on the 1-10 grading scale at the end of a course, with a standard deviation of 1.80. The pre-assignment GPA has a comparable 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.

Panel C summarizes the overall sample structure. We observe 9,825 students interacting with 8,542 peers in 716 courses, and 6,351 different sections.[5] We observe more students than peers because we require peers to show up in at least two distinct sections in the first study year

[5] 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.

and two distinct sections over the rest of their studies. On average, a section has 12.6 students. Through the repeated assignment, a peer interacts on average with 107 students. These interactions are not necessarily unique: conditional on meeting once, two students meet on average 1.3 times. Overall, we observe 916,842 student-peer interactions.

3.3 Randomization Check

The random assignment of students to sections alleviates concerns about the potential non-random selection of students into peer groups. To confirm the randomness of section assignment, we regress student pre-treatment characteristics (previous GPA, age[6], gender, and the rank of the student ID – a proxy for student tenure) on section dummies. This test was first proposed by Wang (2010) and has recently been used by Cullen et al. (2020) to detect ability tracking within schools. Under random assignment, we would expect section dummies to *not* jointly predict pre-treatment characteristics. To check if 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 A1 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.

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 5 percent of the cases and significant at the 1 percent level in approximately 1 percent of the cases. This result confirms that section assignment is stratified based on nationality.

[6] Age is missing for about 20 percent of the sample. For these observations, we impute it using the sample mean.

4. Estimation of Peer Value-Added

Our goal is to obtain an individual-specific estimate for the total contribution of a peer to the performance of others. To do so, we first need to isolate the contribution of peers to performance from common shocks, that is, influences attributed to the classroom environment. Second, we need to isolate the contributions of one peer from other peers in a section. Intuitively, we achieve these goals by exploiting that every student is repeatedly assigned to a unique set of peers. This allows us to compare the average performance of those students who met a specific peer with those who did not.

We take our conceptual framework outlined in Section 2 to the data by estimating student performance as a function of high-dimensional fixed effects at the student-level and peer-level – the workhorse methodology by Abowd et al. (1999, *AKM* hereafter). These peer fixed effects provide our estimates for peer-level value-added. We estimate the following equation using ordinary least squares (OLS):

$$\mathrm{Y}_{\mathrm{ijcs}} = \alpha_{\mathrm{i}} + \sum_{j=1}^{\mathrm{J}} \beta_{\mathrm{j}} \mathrm{D}_{\mathrm{j}} + \delta_{\mathrm{c}} + \nu_{\mathrm{is}} \quad \text{with} \quad \nu_{\mathrm{is}} = \mu_{\mathrm{s}} + \epsilon_{\mathrm{is}}. \tag{3}$$

$\mathrm{Y}_{\mathrm{ijcs}}$ is the course grade of student *i* when interacting with peer *j* in course *c* and section *s*. *J* refers to the total number of peers. Student fixed effects α_i pick up latent and time-invariant student ability. Course-year fixed effects δ_c account for the level of randomization. The error term ν_{is} contains section-level shocks μ_s and idiosyncratic performance fluctuations ϵ_{is}. Given the repeated randomization of students into sections, we can safely assume that the presence of peer *j*, indicated by D_j, is unrelated to any unobserved classroom shocks and idiosyncratic performance fluctuations. Because we exploit random assignment of peers and students to sections, we do not have to worry about section-level shocks in our empirical analysis. This implies:

$$E\left[\nu_{is} \middle| D_j\right] = 0; j = 1, \dots, J. \tag{4}$$

Based on this identifying assumption, estimating β_j with OLS recovers unbiased estimates of the value-added of peer *j*. Accordingly, we denote $\widehat{\beta_j}$ hereafter as $\mathrm{P}\widehat{\mathrm{V}}\mathrm{A}_j$.

Shrinking PVA estimates. One concern with AKM models in worker-firm settings is that worker- and firm-fixed effects are only identified from workers moving between firms and low mobility between firms leads to imprecise estimates (Andrews et al. 2008). This concern is alleviated in our setting as, for every course, all peers are re-assigned to new peer groups. Despite this substantial mobility between peer groups compared to worker-firm settings, random student performance fluctuations between peers remain. These fluctuations will be picked up by $\widehat{\text{PVA}}_j$ and may attenuate the coefficients of $\widehat{\text{PVA}}_j$ when used as an explanatory variable.

To increase the predictive power of the PVA estimates, we follow the teacher value-added literature to shrink consistent but imprecise measures of PVA towards the sample mean (Jacob and Lefgren 2008; Kane and Staiger 2008, Koedel et al. 2015; Chetty et al. 2014a,b).[7] Using empirical Bayes estimates of value-added as independent variables mitigates attenuation bias that would result from using un-shrunken measures in OLS regressions. To construct the empirical Bayes value-added estimates, we multiply the raw $\widehat{PVA}_j$ estimates with an estimate of their reliability. The Bayes estimator shrinks $\widehat{PVA}_j$ towards zero, the prior of PVA:

$$\widehat{PVA}_{j,shrunk} = \widehat{PVA}_j \frac{\widehat{\sigma}^2}{\widehat{\sigma}^2 + \widehat{\sigma}_j^2} \qquad (5)$$

In equation (5), $\widehat{\sigma}^2$ is the total sample variance of the estimated PVA across all observed peers, $\widehat{\sigma}^2 = \frac{1}{n}\sum_j \widehat{\text{PVA}}_j^2$. The average $\widehat{\text{PVA}}_j$ is zero by construction. The term $\widehat{\sigma}_j^2$ describes the within-peer variance of $\widehat{PVA}_j$ across separate teaching-period-specific PVA measures. Intuitively, a peer's $\widehat{PVA}_j$ receives a lower weight when it is based on fewer observations and thus less reliable.

We use the shrunken estimates $\widehat{PVA}_{j,shrunk}$ defined in equation (5) whenever PVA is used as the independent variable. When using peer value-added estimates as a dependent variable, we instead use the raw $\widehat{PVA}_j$ defined in equation (4). Figure A2 in the Appendix shows a strong correlation between raw and shrunken estimates of peer value-added.

[7] Our main results replicate in sign and significance without applying the Bayesian shrinkage.

Common Shocks and the Reflection problem: Two common empirical problems arise in the estimation of standard peer effect models. First, the influence of peers has to be separated from unobservable influences that coincide with the assignment to a peer group, e.g. influences of teacher quality. Such influences have been labeled *common shocks.* The spillovers we measure with PVA are robust against the problem of common shocks. As we observe students in different randomly assigned groups, common shocks, i.e., any kind of unobserved influence that comes with the assignment of peers, will be uncorrelated across the different sections where we observe a peer. We can therefore separate the influence of peers from these idiosyncratic influences which are part of the group-specific environment μ_s and which do not systematically vary with PVA. This intuition is reflected in our identifying assumption expressed in equation (4).

A second concern in the estimation of peer effects is the reflection problem: When testing for contemporaneous effects of peer performance on own performance, it is not possible to disentangle the effect of the peer group on the student from the effect of the student on the peer group. To circumvent this simultaneity problem, the existing peer effects literature has converged to using measures of ability that were determined before the assignment of students to peer groups. The reflection problem does not apply to the estimation of PVA. *At no stage* does our PVA approach rely on regressing student outcomes on peer outcomes. Instead, we estimate a reduced-form effect of the presence of a peer. Even if peer effects would arise through simultaneous effects of a peer's performance on own performance, PVA will remain an unbiased estimate of the total contribution of an individual peer to student performance.[8]

5. Results

In the following section, we present our estimates of PVA. We describe how much peers differ in their PVA in Section 5.1. We then validate our estimates as meaningful measures of peer spillovers in Section 5.2 by establishing that PVA predicts spillovers in out-of-sample social interactions and quantify the importance of an overall pool of peers – all peers a student is exposed to in a course or during all their studies. In Section 5.3, we gain insights into what PVA captures by estimating the relationship between PVA across non-overlapping subsamples of courses and students.

[8] In Appendix 2 we further show that PVA remains a significant predictor in a leave-one-out cross-validation exercise when PVA is based on a sample where individuals enter the estimation *either* as receiving student *or* as affecting peer, but never as both.

5.1 Variation in Peer Value-Added

Peers differ substantially in how much they influence the performance of other students. Figure 1 shows the peer level distribution of PVA. The grey histogram shows the distribution of raw PVA measures obtained from estimating equation (3). The solid red line marks one-standard-deviation of raw PVA. The transparent histogram shows the distribution of PVA after shrinking raw PVA using the method shown in equation (5). On average, PVA measures are shrunk by 73 percent toward zero. The dashed vertical red line shows one standard deviation of shrunken PVA, which is equal to 3.7 percent of a standard deviation in grades. This implies that meeting one peer in one section who has a one standard deviation higher PVA predicts an increase in course grade by 0.037 standard deviations compared to meeting an average peer.

The small average peer value-added masks substantial heterogeneity. Only 16 percent of all peers exert a positive or negative influence larger than 5 percent of a standard deviation in grades. One key advantage of our value-added approach is that we can identify influential peers – bad apples or shining lights – directly from the data. Canonical peer effects models that estimate effects at the group instead of the individual peer level miss these important individual differences.[9]

5.2 Cross-Validation of Peer Value-Added and Peer Capital

Leave-one-out cross-validation. To establish the out-of-sample predictive power of PVA, we construct a jackknife leave-out-measure of PVA. More specifically, for each section we assign peers the average value-added estimated based on all *other* sections, excluding the current one.[10]

---

[9] Hoxby and Weingarth (2005) discuss several competing peer effects models, and investigate whether more flexible models can do a better job in explaining spillovers. Their results highlight the difficulty in modeling peer effects exclusively based on observables, and that the results are dependent on the structure imposed upon peer observable characteristics. More recently, Tincani (2018) relies on flexible semi-parametric methods to estimate heterogeneous effects at any moment of the distribution of observable peer characteristics, which would allow to identify peer effects of bad apples and shining lights.

[10] In practice, we construct leave-one-out jackknife measures as $\widehat{PVA}_{js,jackknife} = (\sum_j \tilde{Y}_{ijcs} - \sum_{j,S\neq s} \tilde{Y}_{ijcs})/(N_j - N_s)$. Here, $\tilde{Y}_{ijcs}$ denotes residuals predicted after a regression based on equation (3) to which estimated peer fixed-effects $\widehat{PVA}_j$ are added back. The first difference denotes the sum of residual performance of all students met by peer *j* minus the sum of residual performance of students in the respective seminar *s*. The second difference denotes the number of students met by peer *j* minus the number of students met by peer *j* in section *s*. Different from the raw PVA, the jackknife measure varies on the peer-section ($j \times s$) level. This approach is similar to validation exercises for teacher value-added in Chetty et al. (2014a) and Jacob et al (2010).

In regressions of grades on PVA, using this jackknife leave-out measure of PVA prevents that the same estimation errors enter both the left and right-hand side of the regression which would bias estimates of the impact of peers. This jackknife PVA measure is highly related to our original PVA. Regressing raw on jackknife leave-out PVA yields a coefficient of $\hat{\beta} = .793$. We shrink our jackknife leave-out measures towards the sample mean as specified in equation (5).

Using this shrunken jackknife measure allows us to do a leave-one-out cross-validation of spillovers in student-peer interactions that were excluded from the construction of PVA.[11] Specifically, we run a regression of student grade (on the student-section level) on the average shrunken jackknife PVA in a section controlling for course and student fixed effects. Figure 2 visualizes the regression results of this exercise. The jackknife measure of PVA is a valid predictor of out-of-sample student performance. A one-standard-deviation increase in jackknife PVA predicts a 0.983 standard deviation increase in performance among the randomly assigned students. The coefficient is highly significantly different from zero and not significantly different from one.

Validating our peer value-added measures does not depend on shrinking value-added estimates. Figure A3 shows the relationship between student grades and average jackknife leave-out PVA that is not adjusted by Bayesian shrinkage. Using un-shrunk measures, we observe a highly statistically significant coefficient of .52.

As an alternative approach we can restrict our sample to those student-peer pairs that meet only once, which leaves us with 80 percent of our estimation sample. Figure A4 shows that the cross-validation coefficient is slightly smaller than in the full sample (.8 of a standard deviation) but again not significantly different from one.

Taken together, the leave-one-out cross-validation confirms that PVA captures systematic variation in peer spillovers instead of random fluctuations coming from the finite number of student-peer interactions. This validation is a prerequisite for deriving valid policy recommendations from peer effects estimates (Carrell, Sacerdote and West 2013).

[11] Appendix Table A2 provides a randomization check for this validation exercise and establishes that Jackknife PVA is unrelated to student characteristics.

Peer capital. To quantify the overall importance of all peers we construct students' *peer capital*. While some students get lucky in the random assignment process and draw higher value-added peers, others draw peers with on average lower peer value-added. We define *peer capital* as the average PVA of a group of peers.

We compute peer capital at two levels: First, we construct a student-section-level measure of peer capital by computing the average jackknife PVA (as defined above) at the section level. This measure consists of the average PVA of all peers one student met in a particular section. Second, we construct a student-level measure by computing the average jackknife PVA of all peers one student met *throughout their studies*. It thus contains the average PVA of all peers met by a particular student, while student performances in *any* section visited by this student do not enter the construction of PVA. Constructing peer capital measures based on the jackknife leave-out measure ensures that contemporaneous performance does not enter estimations twice, once as part of the peer capital measure and once in the dependent variable. Thus, this measure allows us to estimate the causal effect of peer capital on student performance.

Figure 3 shows the section- and student-level distributions of peer capital. Vertical lines in panel (a) and (b) mark the quartiles of the distribution. The bottom panels of Figure 3 show how student performance differs by their peer capital quartile. Panel (c) shows that students who were assigned to a section with peers from the highest quartile of the peer capital distribution have, on average, 5 percent of a standard deviation higher grades than students with peer capital in the lowest quartile. Panel (d) shows similar results for all peers that students meet *throughout their studies*. “Lucky” students in the highest quartile of the peer capital distribution have on average 6 percent of a standard deviation higher GPA than the “unlucky” students with peer capital in the lowest quartile. These results highlight the importance of having an overall, good pool of peers.

Shining Lights and Bad apples: We examine how performance is affected by the number of high- and low-value-added peers in the classroom. To do so, we define *shining lights* and *bad apples* as peers whose PVA falls into the highest and lowest decile of the overall jackknife PVA distribution. Sections contain, on average, 1.2 bad apples and 1.2 shining lights. About 12 percent of all sections contain 3 or more shining lights or 3 or more bad apples.

We regress course grades on dummies for the number of shining lights and bad apples in a section. Figure 4 summarizes the regression results. We find that one single shining light alone does not significantly affect student performance. Starting from two, we observe a linear increase in student performance with each additional shining light in a section. It thus seems that shining lights need a sparring partner in the classroom to create meaningful learning spillovers for other students. The effect of an additional bad apple appears fairly linear over the range of available support.

5.3 What does Peer Value-Added Capture?

One downside of value-added is that estimates largely constitute a "black box": They are not informative about underlying mechanisms. We characterize peer value-added by testing which observable characteristics predict PVA, by analyzing correlations between peer value-added measures for different types of social interactions, and by exploring heterogeneous match effects.

Correlates of peer value-added. We start by analyzing which observable peer characteristics predict PVA in grades. We observe several peer demographic variables frequently studied in the literature: pre-assignment GPA, gender, age, and nationality. These characteristics are highly predictive of students' *own* performance and explain 42 percent of the variation in own achievement, yet their predictive power for PVA is close to zero. Table 3 shows that we observe only weak correlations between PVA and student age, gender, and nationality. *F*-Tests of the (joint) significance of peer observable characteristics in predicting PVA reject the null in several cases. However, the share of explained variation remains below 2 percent in all specifications. Spillovers measured by PVA do not appear to be well captured by observable characteristics commonly used in the peer effects literature. This fact is shared with the teacher value-added literature, where observable characteristics explain little heterogeneity in teacher effectiveness.

Most surprisingly, we do not observe a significant correlation between PVA and pre-determined GPA – our best measure for student ability. Given the strong focus of the peer literature on peer ability as the driving force behind peer effects (Sacerdote 2011), this fact deserves additional discussion. A one-standard-deviation higher pre-assignment GPA predicts a 0.02 percent of a standard deviation higher PVA, reflecting a tiny and economically insignificant relationship. Figure A5 in the Appendix provides evidence that this insignificant relationship is not an artifact of a misspecified non-linear relationship. It is not that higher performers bring out higher performance in others, as implicitly assumed in many models of ability peer effects.

However, our data only comprises a handful of observable student characteristics. We cannot rule out that PVA is systematically related to other characteristics that are unobservable in our data.

The observable characteristics available in our data also fail to meaningfully predict who is in the top and bottom decile of the PVA distribution. In the regression of a bad apple dummy on student GPA, age, gender and nationality, we see that Dutch and German students are more likely to be a bad apple compared to students of other nationalities, while younger students are less likely. Although these coefficients are statistically significant, the R-squared of this regression is only 0.003. In the regression of a shining light dummy on the same independent variables, we see that Dutch students and German students are significantly less likely to be shining lights. However, this regression also only has a tiny R-squared of 0.017.

Consistency of Peer Value-Added across Subjects and Student Subgroups. We construct separate PVA measures based on non-overlapping samples to examine the consistency of PVA. We first estimate separate PVA measures across math-intensive and non-math-intensive courses. This exercise tells us whether the ability to raise others' performance is a general or a subject-specific skill. Panel (a) of Figure 5 shows the relationship between PVA in mathematical versus non-mathematical courses. Regressing un-shrunken PVA in math-intensive subjects on shrunken PVA in non-math-intensive subjects yields a coefficient of $\hat{\beta} = 0.065$, which is significant at the 1 percent level. Thus, PVA appears to be to some degree subject-specific: being a valuable peer in math-intensive courses is only a moderate predictor of being a valuable peer in non-math intensive courses.

Next, we compare how stable PVA is across different student groups. Put differently, we ask if the spillovers one peer creates are the same independent of which kind of students they interact with. To answer this question, we compute PVA based on non-overlapping gender, ability, and nationality groups. For example, we compute the value-added of a given peer on male students, and compare it with the value-added of the same peer on female students. Panel (b) of Figure 5 shows how PVA for male students is correlated with PVA for female students, irrespective of the peer's own gender. The relationship of PVA across gender is substantially larger than the relationship between subjects. Regressing the un-shrunken PVA for male students on the shrunken PVA for female students yields a coefficient of $\hat{\beta} = 0.327$, which is significant at the 1 percent level. Next, we ask if peers affect high- and low-achieving students similarly. As for gender, we compute PVA based on distinct subsamples – separately for below-median-GPA students and above-median-GPA students. Panel (c) of Figure 5 highlights a substantial relationship between PVA measures constructed for low- and high-ability students, similar in magnitude to the relationship across gender. Regressing the un-shrunken PVA for above-median-ability students on the shrunken PVA for below-median-ability students yields a coefficient of $\hat{\beta} = 0.289$, which is significant at the 1 percent level.

Finally, we ask if student nationality plays a role in spillovers captured by PVA. To answer this question, we compute separate PVA measures for student-peer pairs of the same nationality (e.g. both Dutch) and for pairs with different nationalities. Panel (d) of Figure 5 shows the relationship between these two measures, which is weaker than across gender and ability. Regressing the un-shrunken PVA for different-nationality students on the shrunken PVA of same-nationality students yields a coefficient $\hat{\beta} = 0.089$, which is significant at the 1 percent level. One explanation for this weak relationship could be language barriers. While all teaching is held in English, students may find it easier to interact in their mother tongue when studying together outside the classroom, thus reducing across-nationality spillovers.

Taken together, Figure 5 shows that peer value-added measures have substantial overlap across settings (subjects) and characteristics of student-peer matches. Our results suggest that these different PVA measures have a common core that is predictive of spillovers independent of setting or student-peer match. Yet, the relationship is not the same across settings and subgroups. PVA also appears to have a substantial transitory component: social spillovers vary depending on the course content and type of students.

Heterogeneity across the ability distribution. The existing peer effects literature has shown heterogeneous effects along students' own and peers' ability (Carrell et al. 2013, Hoxby and Weingarth 2005, Sacerdote 2014). We test if there are also "match" effects for PVA by exploring if the effect of having a higher value-added peer depends on students' own ability and the magnitude of PVA.

We assess if there are such match effects with two empirical exercises. First, we re-estimate our model allowing for match effects by interacting peer and student fixed effects. Intuitively, we thereby test if the interaction of peer *j* and student *s* leads to interaction-specific spillovers on top of PVA. This specification imposes further restrictions on the data: match effects, student and peer effects, are only separately identified if each particular student-peer combination is observed in at least two sections. This strong, additional restriction reduces the sample by about 45 percent. Table A4 compares model fit and descriptive statistics of peer effects resulting from models with and without interacted peer and student fixed effects (match effects) in the restricted estimation sample. The two alternative PVA measures are strongly related. In the same restricted sample, a simple regression of the original PVA on the match PVA, we get a regression coefficient close to one ($\hat{\beta} = 1.12$, significantly different from zero with p-value < 0.001). Allowing for match effects increases the R^2 by 9 percent and slightly decreases the variance of PVA.

Second, we follow a method introduced by Card, Heining, and Kline (2013) and visually inspect residuals across different types of students and peers. In our context, imposing additivity of peer and student effects could lead to a specification error that results in particularly large or small residuals for certain student-peer matches. For example, if low-achieving students benefit disproportionally from high-value-added peers, we would expect to see particularly large residuals for this combination of students and peers.

Figure A6 in the Appendix shows mean residuals for different combinations of student fixed effects deciles – which capture their ability – and PVA (peer fixed effects) deciles. While mean residuals are small and close to zero in the center of this bivariate distribution, some systematic patterns arise at the tails. Low-ability students appear to perform disproportionally worse when grouped with low PVA peers and benefit disproportionally from high PVA peers. We also see that high-ability students are harmed less by low value-added peers and also benefit less from high PVA peers. These match effects are strongest at the very tails of the respective distributions and moderate in the center.

Taken together, our results suggest that match effects contribute to the variation in performance. Yet, the additive PVA model remains a reasonable specification choice, especially outside the extremes of the student and peer ability distributions.

6. Discussion

6.1 Relationship with the Teacher Value-Added Literature

Figure 6 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 principle value-added is 0.12

standard deviations. Both values are substantially larger than the 0.03 standard deviation variation in PVA we show in this paper. In expectation, replacing a single teacher will have a larger effect on performance than replacing a single peer. Similar to findings in the teacher value-added literature, observable peer characteristics explain little variation in PVA (Chetty et al. 2014).

6.2 Relationship with Prior Work on Peer Effects

Our findings have several implications for how to interpret existing results from the peer effects literature. First, the lack of correlation between PVA and observable characteristics implies that peer effect studies relying on observables such as achievement, gender, or ethnicity do not capture the relevant dimensions in which students affect their peers. Many studies focus on peer effects in one observable dimension and often find small or no peer effects. Such studies might be misleading about the importance of peers.

Second, our paper complements 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.[12] Closest to our paper, three recent studies estimate individual-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. Lacking information on individual player productivity, however, they cannot rely on direct player-peer pair observations, and 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 where 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

[12] 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.

differ in terms of IQ or personality, but score highly on a measure of social intelligence, providing insights into the black box of individual contributions to team production. 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 where peers affect individual performance, as it is typical for educational settings.

Third, peer value-added measures can also be used to (re-)quantify worker contributions to coworkers' productivity. When peer spillovers are substantial, worker-specific performance measures insufficiently describe one's actual contribution to team production. Incentives tied to individual performance may even discourage helping peers to do well and reduce overall productivity. The measures of PVA proposed in the paper could be used to reveal valuable co-workers and quantify their contribution to others' output. PVA could be used to adjust worker-specific productivity measures for contributions that should be attributed to specific co-workers. Knowing who the valuable co-workers are could help to allocate workers into output-maximizing teams. The approach we developed in this paper could also facilitate research on whether workers sort into firms based on co-worker value-added and whether firms reward workers who add more values to their work colleagues.

7. 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 isolate the value-added of a specific peer. In contrast to existing approaches to estimate peer effects, peer value-added summarizes the total contribution to the performance of others capturing both observable and unobservable characteristics that create spillovers. The peer value-added approach does not require taking a stand about which peer observables create spillovers.

We find significant variation in how much peers affect university performance and establish that peer value-added is a valid predictor of spillovers in a leave-one-out cross-validation. The majority of peers, however, have only a small impact on performance: out of all peers with whom students interact, only 16 percent affect grades by more than 5 percent of a standard deviation. We further quantify the importance of an overall pool of peers that a student interacts with during their studies: the *peer capital.* A student with peer capital from the top quartile of the distribution has on average 6 percent of a standard deviation higher GPA than a student with peer capital in the bottom quartile of the distribution. We also show that peers at the extremes of the PVA distributions – bad apples and shining lights – have a disproportional effect. Consistent with the idea that outstanding peers need at least one sparring partner in class, we find that one "shining light" alone does not create any positive spillovers.

We run a series of tests to provide insights into the mechanisms behind PVA. We find that peer value-added is largely unrelated to peers' observable characteristics. Most notably, peer value-added is not correlated with past achievement, implying that higher performers do not necessarily bring out higher performance in others – an implicit assumption in many models of ability peer effects. Finally, we show that peer value-added measures are substantially related across subjects and characteristics of student-peer matches, indicating a common core of PVA that is predictive of spillovers independent of setting or student-peer match.

Our conceptual framework and estimation provide a new perspective on how to think about peer effects. Peer value-added allows estimating peer effects without priors about which observable peer characteristics generate spillovers. Peer value-added can even be estimated in the absence of information on observable peer characteristics. The peer value-added 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

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

Andrews, M. J., Gill, L., Schank, T., and Upward, R. (2008). High wage workers and low wage firms: negative assortative matching or limited mobility bias?. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3), 673-697.

Angrist, J. D., Hull, P. D., Pathak, P. A., and Walters, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation. *The Quarterly Journal of Economics*, 132(2), 871-919.

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.

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.

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., 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.

Cullen, J. (2020). The causes and consequences of ability tracking: Evidence from Texas public schools. Unpublished manuscript.

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.

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

Hoxby, C. M., and Weingarth, G. (2005). Taking race out of the equation: school reassignment and the structure of peer effects. *NBER Working Paper 7867*.

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.

Jacob, B. A., Lefgren, L., and Sims, D. P. (2010). The persistence of teacher-induced learning. *Journal of Human Resources*, 45(4), 915–943.

Jackson, K., and Bruegmann, E. (2009). Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers. *American Economic Journal: Applied Economics, 1*(4), 85–108.

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

Koedel, C., Mihaly, K., and Rockoff, Jonah 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.

Putnam, R. D. (1995). Bowling Alone: America's Declining Social Capital. *Journal of Democracy*, 6(1), 65–78.

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.

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.

Tincani, M. M. (2018). Heterogeneous peer effects in the classroom. *Working Paper.*

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	9,825	20.01	1.66	15.93	40.95
Female	9,825	0.39	0.49	0	1
Dutch	9,825	0.23	0.42	0	1
German	9,825	0.38	0.48	0	1
B. Student performance					
Pre-assignment GPA	64,171	6.86	1.18	1	10
Course grade	77,457	6.57	1.80	1	10

C. Summary Statistics of Dyadic Student-Peer Interactions	Number
Number of course-year observations	716
Number of sections (peer groups)	6,351
Number of individual students	9,825
Number of individual peers	8,542
Number of student-course observations	77,457
Average number of students per section	12.57
Total number of dyadic student-peer interactions	916,842
Average number of student-peer interactions by peer	107.33
Repeated student-peer interactions	1.28

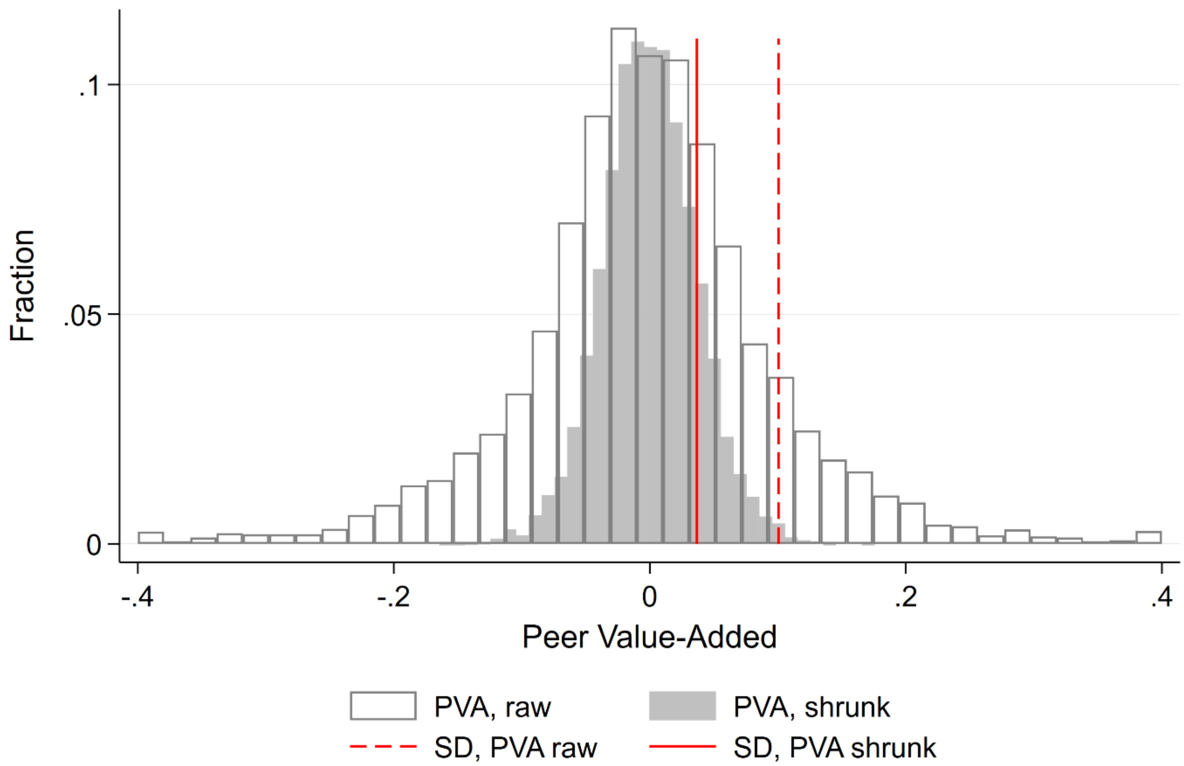
Note: This table shows descriptive statistics of the estimation sample. 'SD' refers to the standard deviation of the respective variable. Panel A reports individual student demographic characteristics. Panels B reports measures of student performance: pre-section assignment grade point average (GPA) and end-of-course grades at the student-course level. The number of individual students and peers differs because we have dropped peers observed in fewer than 4 sections. Pre-assignment GPA in Panel B is missing for students in the first period.

Table 2: Test for Random Assignment of Students to Sections

Dependent variable:	(1) Number of total tests performed	(2) Number significant at the:			(3) Percent significant at the:		
		5%	1%	0.1%	5%	1%	0.1%
Female	679	25	6	1	3.7%	0.9%	0.1%
GPA	610	40	11	1	6.6%	1.8%	0.2%
Age	668	28	4	0	4.2%	0.6%	0.0%
ID rank	680	35	2	0	5.1%	0.3%	0.0%
Dutch	658	18	4	1	2.7%	0.6%	0.2%
German	649	14	3	0	2.2%	0.5%	0.0%

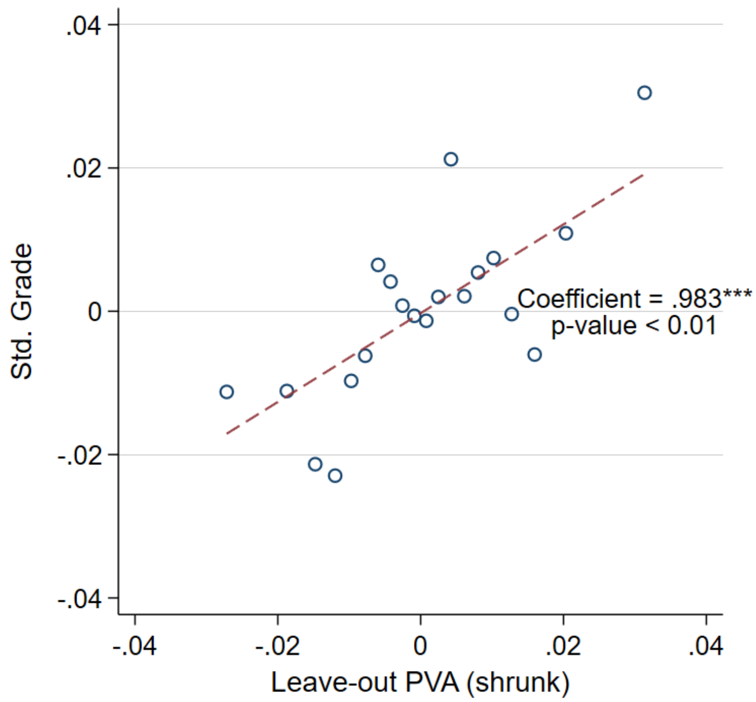
Note: This table is based on separate OLS regressions with gender, GPA, age, and ID rank (a proxy for student tenure at the university) as dependent variables. The explanatory variables are a set of section dummies. Columns (1) and (2) show in how many regressions the *F*-statistic of joint significance 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 A1 shows histograms the *p*-values of each performed test from the underlying regressions.

Figure 1: Variation in Peer Value-Added



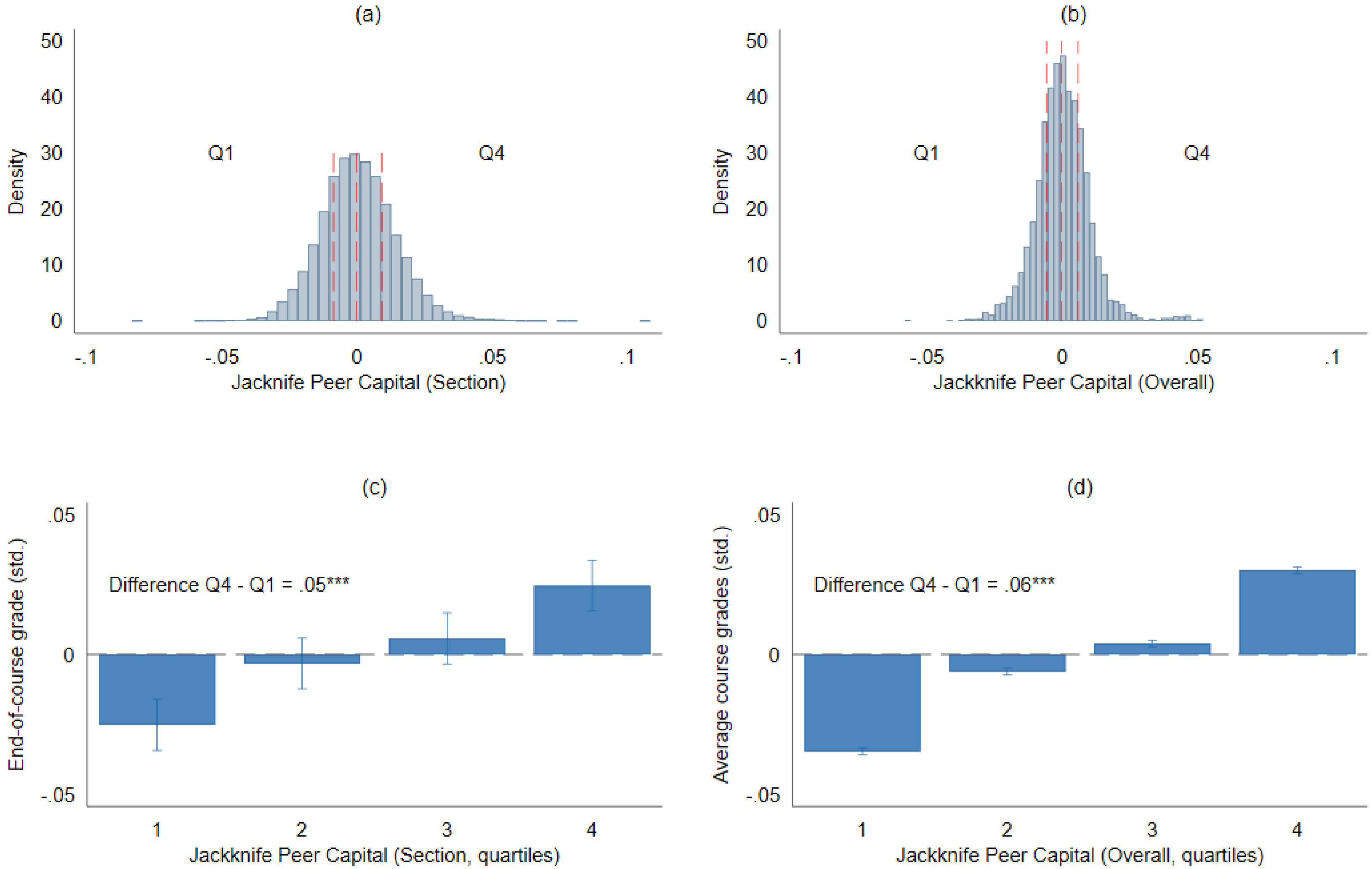
Note: This figure shows the value-added estimates at the level of 8,542 peers. Underlying grades are standardized to mean zero and unit variance. The dashed line shows raw PVAs estimated as peer-level fixed effects explaining grades, and conditioning on student-level and course-level fixed effects (equation 1). The grey bars display the related distribution of shrunken PVA adjusted by Bayesian shrinkage. Vertical red lines mark standard deviations: SD (raw PVA, dotted line) = 0.10, SD (shrunken PVA, dashed line) = 0.037. Estimations are based on 916,842 student-peer interactions.

Figure 2: Jackknife Validation of Peer Value-Added



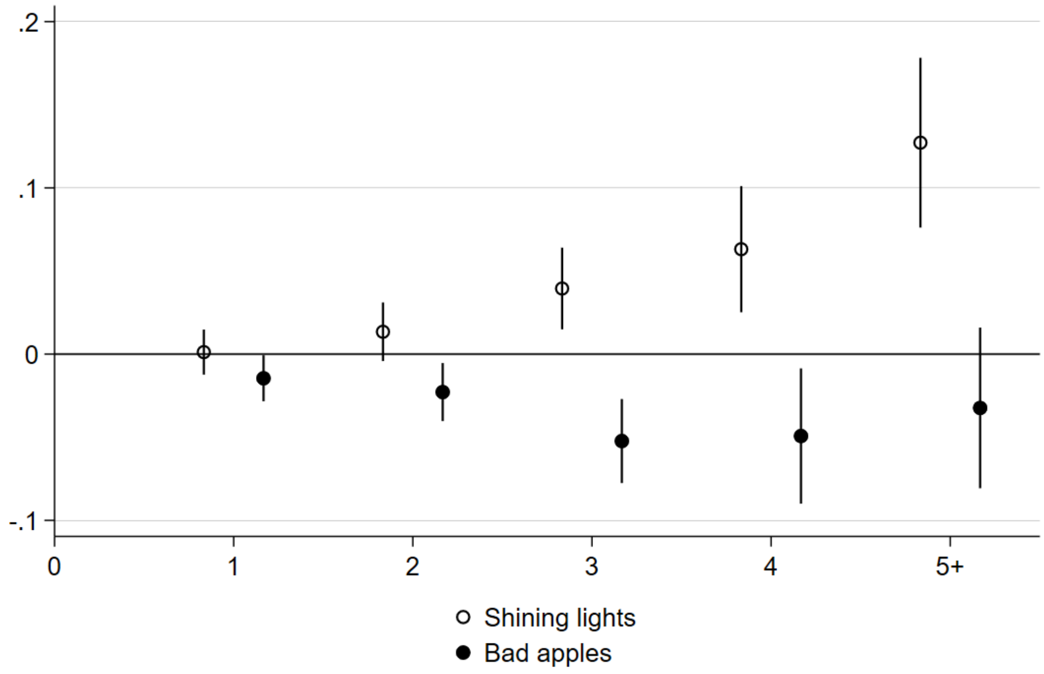
Note: This binned scatterplot visualizes the relationship between student grades and average shrunken jackknife leave-out PVA which is constructed based on a section-leave-out sample that excludes the section where we observe the grade we use as the dependent variable. The point estimate and significance level are obtained from a regression of grade on shrunken leave-out PVA, and conditioning on student-level and course-level fixed effects. Both grade and the shrunken PVA are standardized to have means of zero and standard deviations of one for the estimation sample. One observation is one student-section observation. N= 77,457. All p-values care based on two-way clustering at the student and section level. *** indicate significance level p<0.01.

Figure 3: Peer Capital



Note: This figure describes differences in student performance by peer capital. Peer capital is computed as the average of jackknife PVA at the student-section level and the overall student level, respectively. Panel (a) shows the distribution of peer capital at the section (N= 77,457) level. Panel (b) shows the distribution of peer capital at the student level (N= 9,825). Red dashed lines indicate quartiles. Panels (c) and (d) show how student grades and GPA differ by quartiles of students' peer capital. Panel (c) is based on a regression of course grades on peer capital quartiles. Panel (d) is based on a regression of student GPA on peer capital quartiles. *** indicate significance level p<0.01. Error bars indicate 95% confidence interval.

Figure 4: Bad Apples and Shining Lights



Note: This figure shows point estimates of a regression of standardized student grades on the number of high- and low-value added peers at the student-section level. All coefficients are estimated simultaneously. N= 77,457. High value-added peers ("shining lights") are defined as peers from the top decile of the PVA distribution, based on jackknife PVA. Low value-added peers ("bad apples") are defined as peers from the bottom decile of the jackknife PVA distribution. Vertical lines indicate 95% confidence intervals. Confidence intervals are based on robust standard errors clustered at the section level.

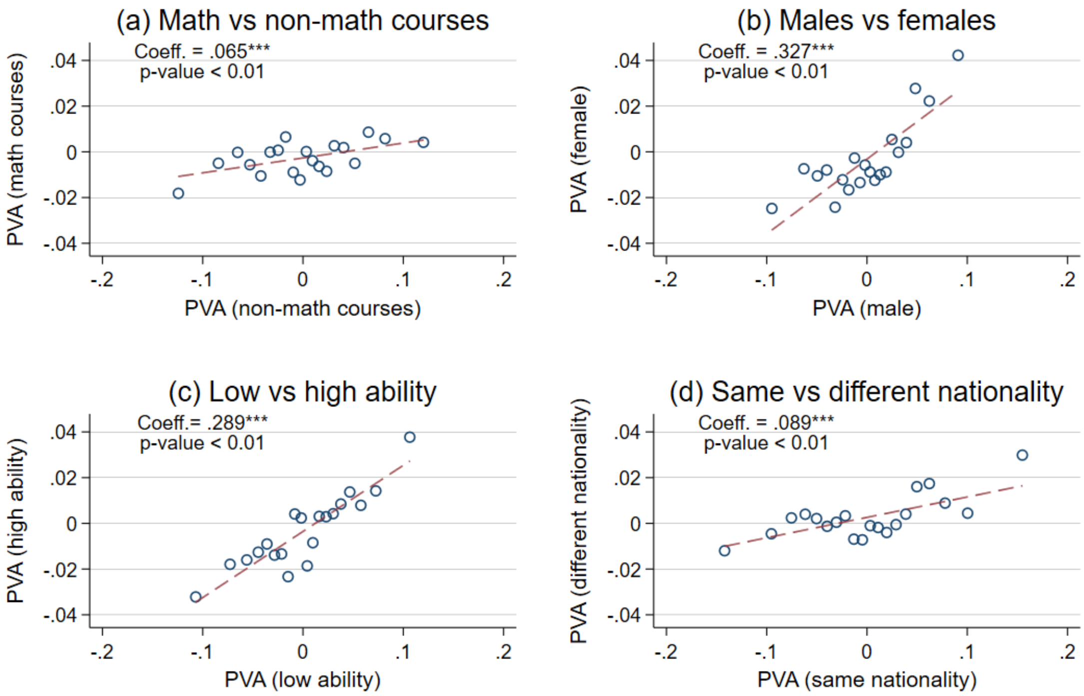
Table 3: Who is a Good Peer?

Correlation between Student Characteristics and Peer Value-Added

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable:	PVA	PVA	PVA	PVA	PVA	PVA	Top decile of PVA "shining light"	Bottom Decile of PVA "bad apple"
GPA (std.)	0.0002					0.0000	0.0062*	0.0056*
	(0.0011)					(0.0011)	(0.0037)	(0.0032)
Female		0.0068***				0.0076***	-0.0094	-0.0102
		(0.0022)				(0.0023)	(0.0069)	(0.0069)
Age			0.0020***			0.0027***	0.0028	-0.0040**
			(0.0007)			(0.0008)	(0.0024)	(0.0020)
Nationality: Dutch				0.0003		0.0033	-0.0796***	0.0173**
				(0.0022)		(0.0030)	(0.0092)	(0.0088)
Nationality: German					0.0002	-0.0009	-0.0842***	0.0320***
					(0.0021)	(0.0028)	(0.0085)	(0.0077)
Observations	7,980	8,194	8,194	8,194	8,194	7,980	7,980	7,980
R-squared	0.0000	0.0011	0.0012	0.0000	0.0000	0.0031	0.0173	0.0030
Prob > F	.8737	.0025	.0039	.8803	.9109	.0007	<.0001	.0001

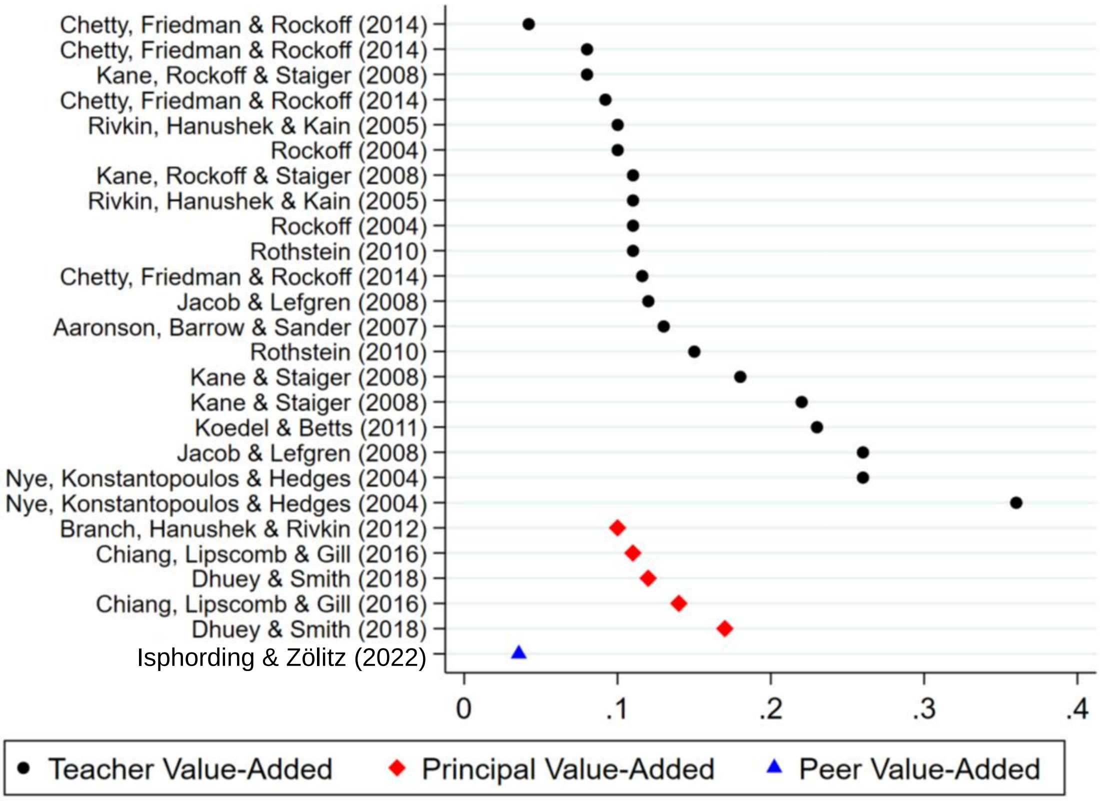
Note: This table reports OLS regression results of raw PVA on peer observable characteristics with one observation per peer. GPA is standardized to have a mean of zero and standard deviation of one for our estimation sample. The dependent variables in columns (7) and (8) are indicators for being in the 1st or 10th decile of the PVA distribution, respectively. Smaller numbers of observations in columns (1), (6), (7), and (8) due to missing pre-assignment GPA in the first assignment period. Robust standard errors are in parentheses. Significance levels indicated as * p<0.1, ** p<0.05, and *** p<0.01.

Figure 5: Correlation of Peer Value-Added across Subjects and Student Subgroups



Note: These bin scatter plots show how a peer's value-added varies across settings and groups of students. The upper left panel shows the relationship between a peer's value-added in math- and non-math-intensive courses. The upper right panel shows the relationship between a peer's value-added on male and female students. The lower left panel shows the relationship between a peer's value-added on above-median (high) and below-median (low) ability students. The lower right panel shows the relationship between a peer's value-added in student-peer interactions with same and different nationality Point estimates and significance levels are obtained from pairwise OLS regression on the peer level. All *p*-values are based on heteroskedasticity robust standard errors. *** indicate significance level p<0.01.

Figure 6: 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), and peers (blue triangle) in recently published studies. All estimates refer to standard deviations of value-added distributions constructed for standardized performance measures.

ONLINE APPENDIX

APPENDIX A1: Additional Tables and Figures

Table A1: Data Structure before and after Reshaping

Panel A: Original data before reshaping		
Section A	Anne	
Section A	Dick	
Section A	Julian	
Section B	Anne	
Section B	George	
Section B	Timmy	
Number of observations	6	
Panel B: Dyadic data after reshaping		
Section A	Anne	Julian
Section A	Anne	Dick
Section A	Dick	Anne
Section A	Dick	Julian
Section A	Julian	Anne
Section A	Julian	Dick
Section B	Anne	George
Section B	Anne	Timmy
Section B	George	Anne
Section B	George	Timmy
Section B	Timmy	Anne
Section B	Timmy	George
Number of observations	12	

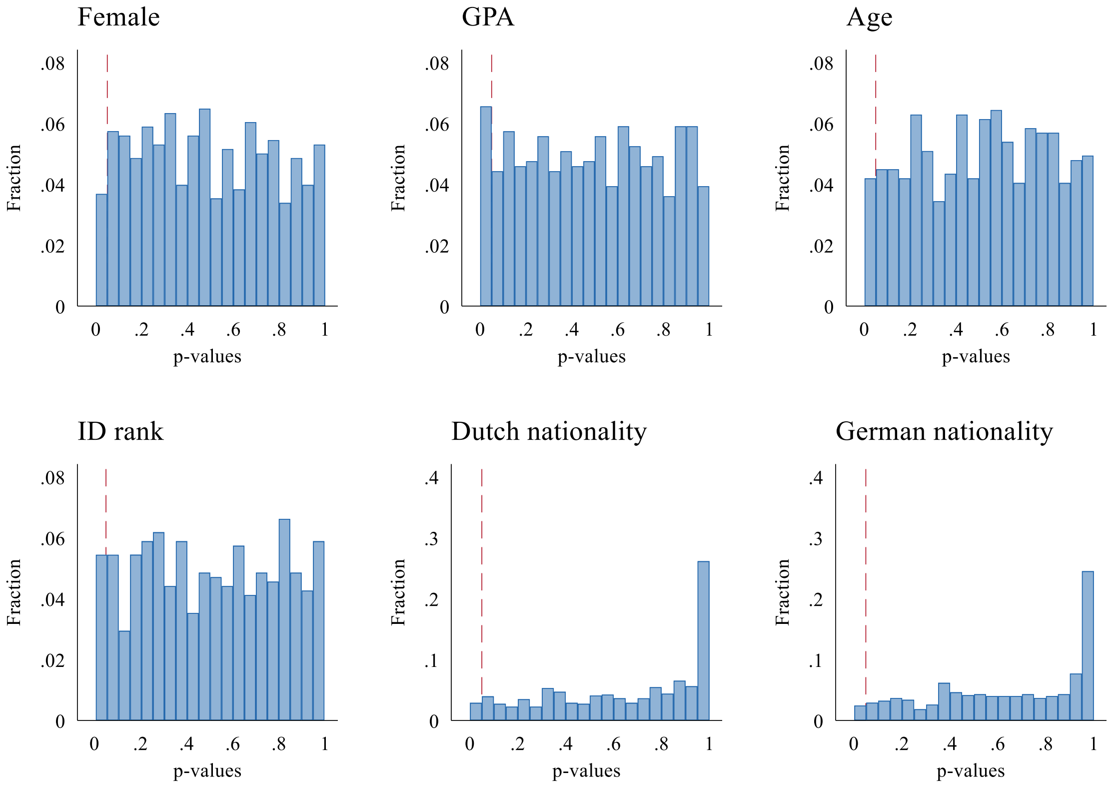
Note: In Panel A each observation represents one student-class observation. In Panel B, each observation represents one student-peer interaction. When reshaping the data, the number of observations increases from $\sum_{c=1}^{C}\sum_{s=1}^{S} n_{cs}$ to $\sum_{c=1}^{C}\sum_{s=1}^{S} n_{cs}\,(n_{cs}-1)$.

Table A2: Randomization Check for Leave-out Jackknife Validation

	(1) Mean PVA in section	(2) Mean PVA in section	(3) Mean PVA in section	(4) Mean PVA in section	(5) Mean PVA in section	(6) Mean PVA in section
GPA	-0.0000					-0.0000
	(0.0001)					(0.0001)
Age		0.0000				-0.0000
		(0.0001)				(0.0001)
Female			0.0002			0.0002
			(0.0002)			(0.0002)
German				0.0000		-0.0001
				(0.0001)		(0.0002)
Dutch					-0.0000	-0.0001
					(0.0002)	(0.0002)
Observations	64,171	77,457	77,457	77,457	77,457	64,171
R-squared	0.3202	0.3149	0.3149	0.3149	0.3149	0.3202

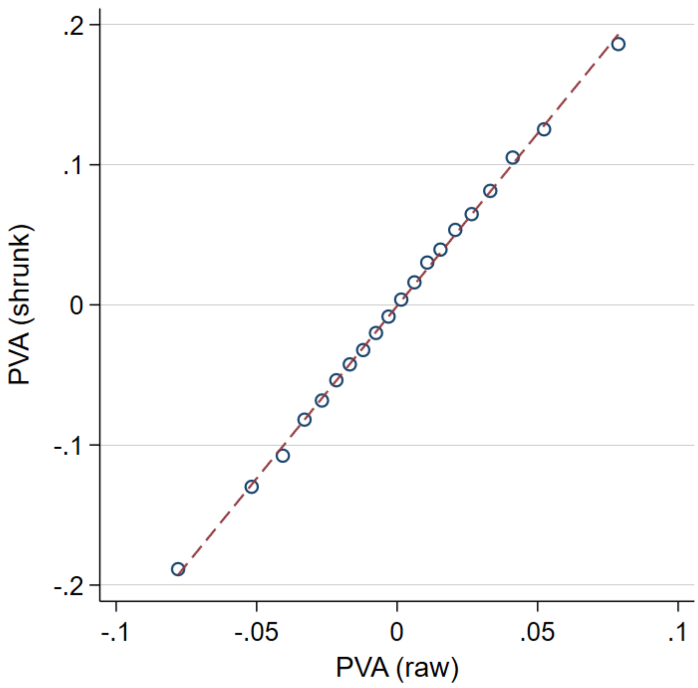
Note: This table reports results of regressions of mean jackknife leave-out PVA by section on pre-determined student characteristics, conditional on course-level fixed effects. One observation is one student-section observation. Robust standard errors using two-way clustering at the individual and section level are in parentheses. Smaller numbers of observations in columns (1) and (6) are due to missing pre-assignment GPA in the first assignment period.

Figure A1: Randomization Check - Distribution of *p*-values



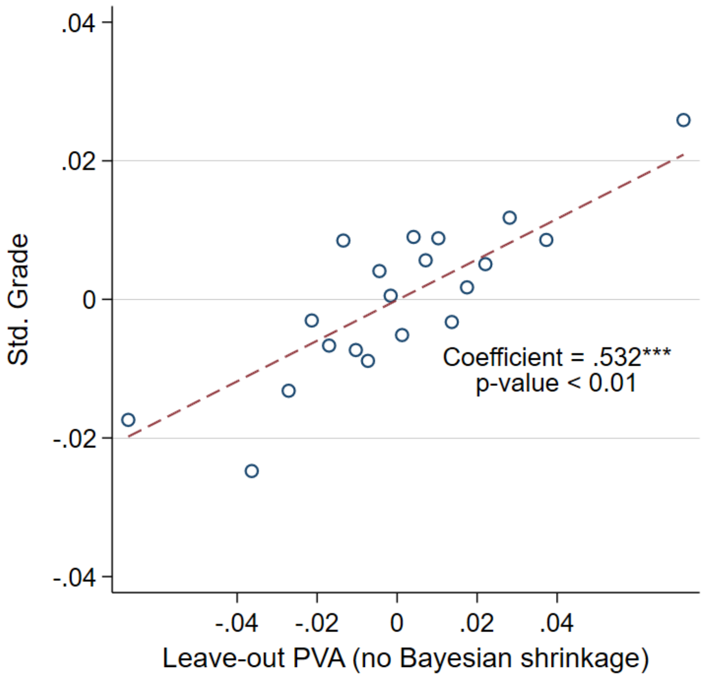
Note: These histograms show *p*-values from all the regressions reported in Table 2. The vertical line in each histogram shows the 5-percent significance level. Dutch and German are mechanically "over-balanced" due to the stratified assignment based on nationality.

Figure A2: Correlation between raw and shrunken PVA



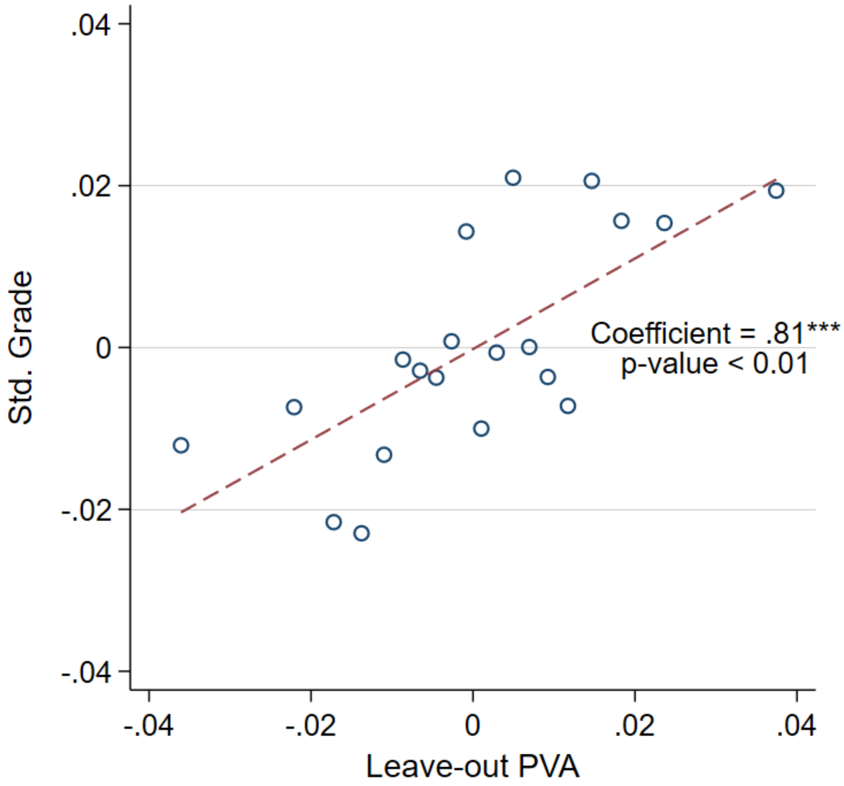
Note: This bin scatter shows the relationship between raw and Bayesian shrinkage-adjusted ("shrunken") PVA. Both measures correlate with r=0.90. N=8,542.

Figure A3: Leave-One-Out Cross-Validation using raw PVA



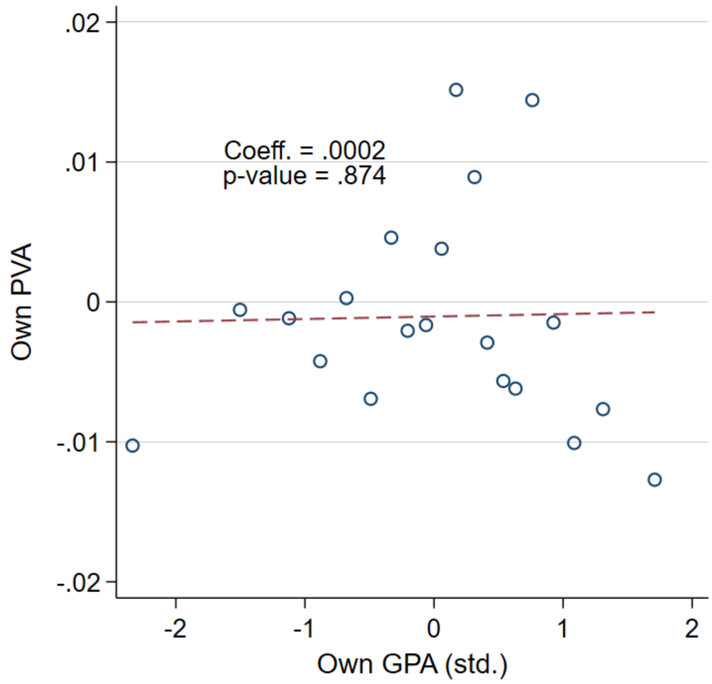
Note: This bin scatter visualizes the relationship between student grades and average jackknife leave-out PVA (not adjusted by Bayesian shrinkage) which is constructed based on a section-leave-out sample that excludes the section where we observe the grade that we use as the dependent variable. The point estimate and significance level are obtained from a regression of standardized to mean zero and unit variance on average shrunken leave-out PVA, and conditioning on student-level and course-level fixed effects. One observation is one student-section observation. N= 77,457. All *p*-values are based on two-way clustering at the student- and section-level. Significance levels indicated as * p<0.1, ** p<0.05, and *** p<0.01.

Figure A4: Leave-One-Out Cross-Validation using only Unique Tuples



Note: This bin scatter visualizes the relationship between student grades and average shrunken jackknife leave-out PVA which is constructed based on a section-leave-out sample that excludes the section where we observe the grade that we use as the dependent variable. The sample is restricted to student-peer interactions of individuals that meet only once during the sample period. The point estimate and significance level are obtained from a regression of standardized to mean zero and unit variance on average shrunken leave-out PVA, and conditioning on student-level and course-level fixed effects. One observation is one student-section observation. N=77,457. All p-values are based on two-way clustering at the student- and section-level. Significance levels indicated as * p<0.1, ** p<0.05, and *** p<0.01.

Figure A5: Correlation between GPA and Peer Value-Added



Note: This binned scatterplot visualizes the relationship between own ability (measured as standardized GPA) and PVA. The point estimate and significance level are obtained from a regression of own raw PVA on own pre-assignment standardized grade point average (GPA). One observation is one student-section observation. N=7,980. All p-values are based on robust standard errors.

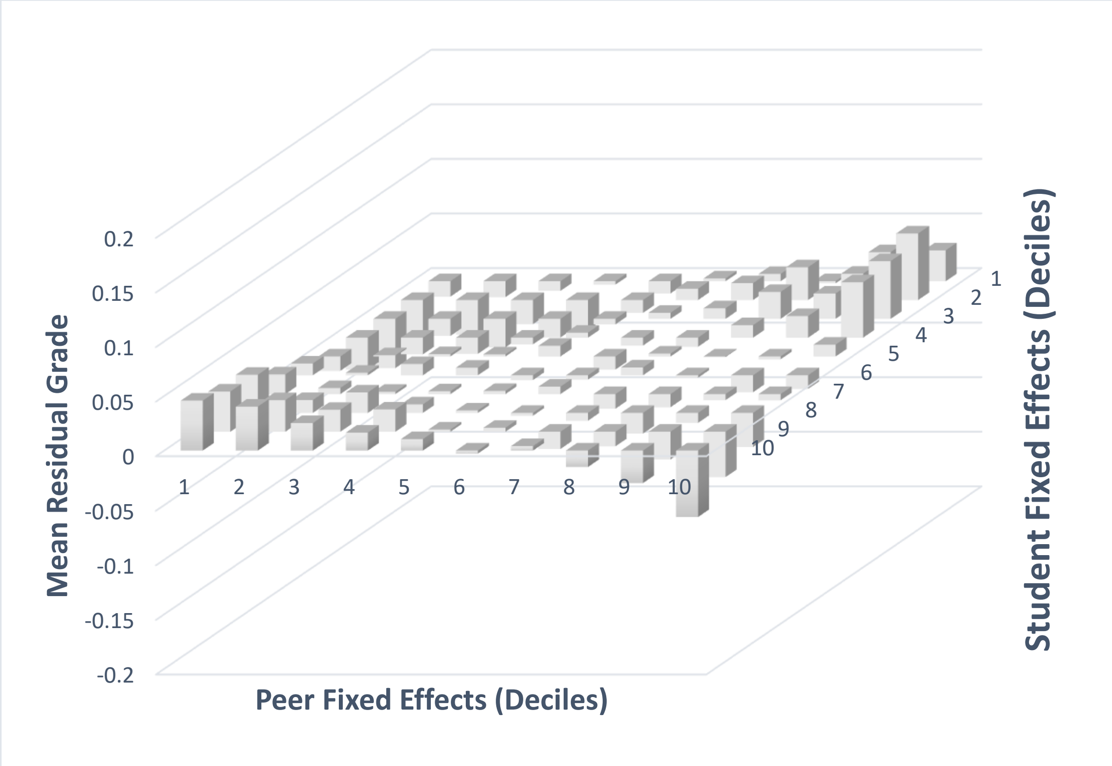
APPENDIX A2: Indirect Reflection

The reflection problem is a general concern in the estimation of peer effects and makes it challenging to disentangle the effect of the peer on the student from the effect of the student on the peer (Manski, 1993). The existing peer-effects literature has converged to using pre-treatment characteristics to circumvent the reflection problem.

The reflection problem does not impede the PVA approach because we do not regress student outcomes on peer outcomes. We nevertheless conduct a simple empirical exercise to show that our results are not an artifact of a simultaneity problem. We conduct the following straightforward test: We estimate PVA in a dataset in which we assign to each individual with an even student ID the *peer* role, and to each individual with an odd ID the *receiving student* role. This way each single student enters the estimation *either* as a peer or student, but never in both roles simultaneously.

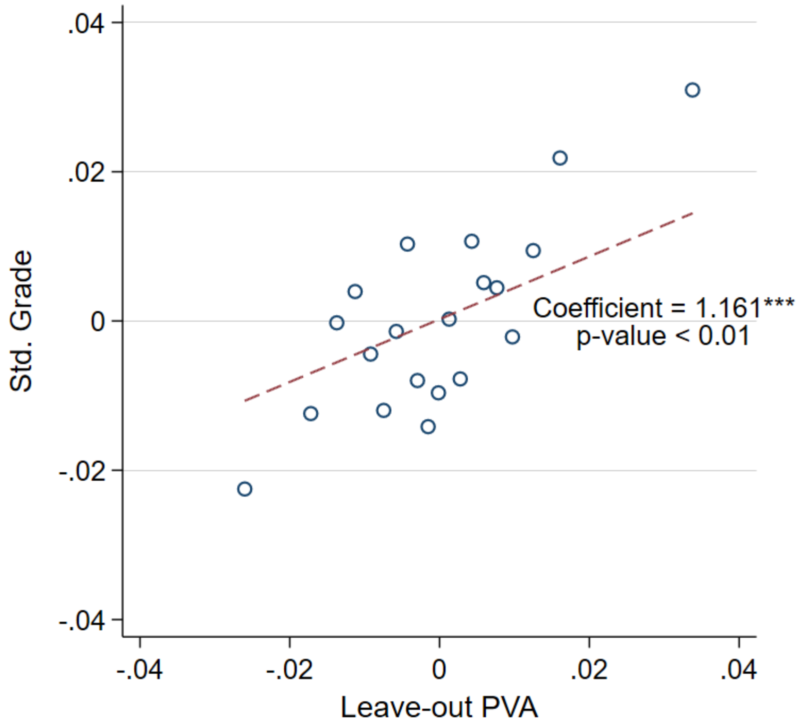
Computing PVA in this way provides results comparable to the main specification. Figure A7 shows the result of a leave-one-out cross-validation exercise that is identical to Figure 2 but based on the distinct peer and student samples as described above. The resulting coefficient is 1.16, significant at the 1 percent level and not significantly different from 1. We are therefore not concerned about simultaneity affecting our PVA estimates.

Figure A6: Average Residuals by Student – and Peer Effect Decile



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 (peer value-added) by deciles of the student- and peer-effects distribution. Bars indicate mean residuals in cells.

Figure A7: Jackknife Validation: PVA based on Disjunct Peer and Student Samples



Note: This bin scatter visualizes the relationship between student grades and average shrunken jackknife leave-out PVA adjusted in the following way: Different from Figure 2, individuals enter the construction of PVA either as ``affected'' students or as ``affecting'' peers, but not both. PVA for peers with even student ids is based on the performance of students with odd student ids only and vice versa. Based on this adjusted PVA, we construct a section-leave-out sample that excludes the section where we observe the grade that we use as the dependent variable. The point estimate and significance level are obtained from a regression of standardized to mean zero and unit variance on average shrunken leave-out PVA, and conditioning on student-level and course-level fixed effects. One observation is one student-section observation. N=76,385. All p-values care based on two-way clustering at the student- and section level. Significance levels indicated as * p<0.1, ** p<0.05, and *** p<0.01.

Table A4: Peer Fixed Effects Estimations with and without Match Effects

	No Match Effects	Match Effects
Peer Fixed Effects (SD)	.111	.075
No. of observations	350,296	350,296
Model Fit (R^2)	0.642	0.774

Note: This table reports summary statistics of regressions of student grades on course-, student- and peer fixed effects. Column (1) shows the results of the additive fixed effects model. Column (2) shows estimation results including interactions between student and peer effects. Reported statistics are the standard deviation of peer fixed effects and the number of observations and the model fit (R^2).