The Value of a Peer*

INGO E. ISPHORDING
(IZA)

ULF ZÖLITZ
(University of Zurich)

March 2020

Abstract

This paper introduces peer value-added, a new approach to quantify the total contribution of an individual peer to student performance. Peer value-added captures social spillovers irrespective of whether they are generated by observable or unobservable peer characteristics. Using data with repeated random assignment to university sections, we find that students significantly differ in their peer value-added. Peer value-added is a good out-of-sample predictor of performance spillovers in newly assigned student-peer pairs. Yet, students' own past performance and other observable characteristics are poor predictors of peer value-added. Peer value-added increases after exposure to better peers, and valuable peers are substitutes for low-quality teachers.

Keywords: peer effects, peer value-added, peer capital, spillovers
JEL classification: I21, I24, J24

* We thank our valuable peers Steffen Altmann, David Dorn, Jan Feld, 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. Jeffrey Yusof and Maximilian Mähr 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

The social nature of humans creates constant spillovers. 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 generate spillovers, 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 a given peer's contribution to others' performance. Peer value-added summarizes *any* source of spillover on others' performance, regardless of whether spillovers stem from observable or unobservable peer characteristics, skills, or behaviors. In fact, peer value-added can even be estimated when information on observable peer characteristics is entirely absent. While peer value-added is a person-specific measure, it also can be aggregated to *peer capital*. Peer capital is a student-level measure of the overall peer quality that an individual student has experienced.

We estimate peer value-added in a business school setting where students are repeatedly and randomly assigned to groups of 10 to 16 peers. Based on data with 383,040 dyadic student-peer pairs, we construct peer value-added as the average residual performance of students who met a specific peer. We shrink these peer-level residuals toward the sample mean to increase its out-of-sample predictive power when used as a regressor.

We document five sets of results. First, peers significantly differ in their ability to improve their fellow students' performance. Meeting a peer with a one-standard-deviation higher value-added increases a student's grade by 3.6 percent of a standard deviation. However, the majority of peers only have small impacts on performance: out of all peers with whom students interact, only 12.5 percent affect performance by more than 5 percent of standard deviation.

Second, observable peer characteristics are poor predictors of peer value-added. Most noteworthy, lagged achievement is only weakly correlated with peer value-added. This result runs contrary to the popular belief that high-achieving students bring out the best in others. While from an econometrician's perspective peer value-added is mostly unrelated to observable characteristics, fellow classmates notice valuable peers. Students report significantly improved peer-to-peer interactions in teaching evaluations when high value-added peers are present in the classroom.

Third, peer value-added is systematically correlated across subjects and time. The same peers who raise performance in mathematical courses also raise performance in non-math-intensive courses. Importantly, peer value-added is a strong predictor for spillovers in newly assigned *out-of-sample* social interactions.

Fourth, peer value-added is malleable and increases after being assigned to classes with more valuable peers in earlier periods. Students who meet high value-added peers not only perform better themselves, but subsequently also produce larger spillovers on others, thus becoming more valuable peers.

Fifth, peer and teacher value-added are substitutes. Peers matter most when the class is taught by a low value-added instructor, showing that high value-added peers can compensate for the adverse effects of less effective teachers.

With this paper, we make both conceptual and empirical contributions to the peer effects literature. Conceptually, by focusing on the person-specific peer value-added, we differ 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. For example, this literature studies how student test scores are affected by having higher-achieving peers, or more black, female or free-lunch peers.[1] In contrast to our framework, these studies are motivated by the idea that the ability of raising others' performance is closely related to one's own observable characteristics.[2] By recognizing that spillovers can arise from both observed and unobserved peer characteristics, we propose a more comprehensive approach to estimate peer effects. This approach does not rely on a prior about which observables can produce peer effects.

Empirically, we contribute to a more nuanced understanding of peer effects in education by providing first-time estimates of the person-specific peer value-added. The poor correlation between peer value-added and observable peer characteristics indicates that social spillovers are not necessarily well captured by student observables. By providing empirical evidence on the stability and malleability of peer value-added, we gain insights into the environmental factors that shape peer spillovers. Finally, while both teacher and peer effects have been intensively studied in isolation, this is the first paper to provide evidence on how peer effects interact with teacher value-added. By showing that peer value-added can act as a compensating substitute for low-quality teachers, we provide new evidence on the interaction of two key parameters in education production.

[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 recent evidence on ability peer effects.[2] Fruehwirth (2014) criticizes the assumption that peer spillovers can be measured through observables, as theories of peer spillovers often focus on unobservables. She further shows that typical empirical specifications often do not capture spillovers accurately.

[2] Fruehwirth (2014) criticizes the assumption that peer spillovers can be measured through observables, as theories of peer spillovers often focus on unobservables. She further shows that typical empirical specifications often do not capture spillovers accurately.

Our approach to identify peer value-added is inspired by the literature on teacher value-added.[3] Similar to that literature, we find that observable characteristics are not very good predictors for spillovers. A comparison of value-added estimates of teachers on standardized test scores to peer value-added on standardized student performance shows that the variation in teacher value-added is three to five times larger than the variation in peer value-added. This suggests that differences between individual teachers affect performance more than differences between individual peers. Peer value-added is also less stable over time, with a year-to-year correlation of about 0.1, compared to a range of 0.2 to 0.7 for teacher value-added. We show that changes in peer value-added can be partly explained by exposure to high value-added peers in earlier periods.[4]

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 the estimation steps. In Section 5, we present results on the magnitude, heterogeneity, and malleability of peer value-added. We further show that peer value-added is correlated across subjects and time, and that it is predictive for spillovers in new "out-of-sample" social interactions. Section 6 discusses the findings in relation to the existing literature and in terms of potential applications. Section 7 concludes.

2. Conceptual Framework

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

[3] Koedel et al. (2015) and Hanushek and Rivkin (2006) provide a comprehensive review on the extensive literature on teacher value-added.
[4] Jackson and Bruegmann (2009) present similar results for teachers, and show that teacher value-added is affected by the value-added of fellow colleagues.

The PVA approach differs from the existing literature on peer effects in two important respects. First, PVA identifies the *person-specific* contribution to performance. It describes the expected change in student performance caused by meeting one specific peer relative to meeting a peer with average value-added. By contrast, the standard peer effects approach focuses on the contribution of an entire *group of peers*. The contribution of this group is, for example, estimated as the effect of peer group averages of test scores or the share of female, black, or free-lunch-receiving peers. This aggregation to group-level variables implies a loss of information. Furthermore, it is not possible to isolate the contribution of an individual peer in such a design.

Second, PVA measures the *total contribution* to others' performance by capturing 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. In fact, PVA can be estimated even when data on student characteristics are entirely absent. This distinction is particularly important when peer effects are driven – for example – by socio-emotional skills or behavioral differences, which are often not observed. In contrast to our approach, standard peer effects studies can only identify spillovers in dimensions that are either observed or sufficiently correlated with observable characteristics. These studies might therefore falsely conclude that peer effects are small or non-existent when observable characteristics do not sufficiently capture spillovers.

We describe the PVA approach in a simple education production function framework that includes the peer environment as a separate factor:

$$y_{is} = f\left(\delta_s^{-i},\ \alpha_i\right) + \epsilon_{is}. \tag{1}$$

In equation (1), y_{is} represents the performance of student i in peer group s, which is assumed to be a function of factors associated with the peer group environment δ_s^{-i}, factors associated with the student α_i, and idiosyncratic variation ϵ_{is} of student performance in a specific section, such as performance fluctuations on the day of an exam.

The peer group environment δ_s^{-i} contains all influences that are related to the composition of a specific peer group, excluding student *i*. The PVA approach aims to decompose the peer environment δ_s^{-i} into its individual peer-specific components:

$$\delta_s^{-i} = \sum_{j \neq i, j \in s} \omega_{ijs} + \mu_s. \tag{2}$$

The term $\sum_{j \neq i, j \in s} \omega_{ijs}$ describes the sum of all spillovers ω_{jis} of peer *j* on student *i* within a specific peer group *s*. Equation (2) implies that peers are perfect substitutes for one another and that peers contribute to others' performance in an additive and separable way. Remaining environmental factors jointly assigned with the peer composition are denoted by μ_s.

Consider now that peer *j* is observed in different peer groups. Equation (2) allows for spillovers of peer *j* to vary over students *i* and peer environments *s*. The person-specific and comprehensive spillover PVA_j is then defined as the average spillover of peer *j* over all N_j student-peer interactions with students $i \neq j$ across all peer groups *s* where peer *j* is present:

$$PVA_j = \frac{1}{N_j} \sum_{i \neq j, i \in s} \omega_{ijs} \tag{3}$$

PVA_j captures the expected change in outcome y_{is} that can be attributed to the presence of the specific peer j. Variation in peer-specific value-added can arise from differences in peer behavior, 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 inspiring effort provision in their fellow students. Furthermore, some peers might possess specific personality traits that generate a productive learning environment for their fellow students. These mechanisms also can be negatively framed: students can be disruptive, possess a non-complementary skill set, discourage other students from providing effort, and generate a hostile learning environment through their personality. Irrespective of the exact source, PVA_j will capture the expected contribution of peer j to the performance of student i.

Defining PVA_j as the average spillover over all student-peer pairs containing j captures a context-independent component of spillovers. However, the ability to raise others' performance might depend to some degree on the specific peer composition, curriculum, or teacher in a classroom. These transitory deviations of student-peer-specific ω_{jis} from PVA_j capture potential interactions of a peer's spillover with contextual factors like teaching styles, curricula or other peers' behavior.

Empirically, a peer-specific spillover is not directly observable in the data, but has to be inferred from performance changes of students who meet different peers. In the empirical part of the paper, we will use repeated observations of students in changing peer groups to isolate PVA_j. We calculate peer value-added by computing the average performance gains of students assigned to a specific peer j relative to meeting peers other than j. We describe our setting and estimation strategy in detail in Sections 3 and 4.

3. Institutional Environment, Data and Randomization

We estimate PVA based on 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.[5]

3.1 Institutional Environment

We use data from a Dutch business school that offers bachelor, master, and PhD programs. Our analysis focuses on two large study programs in which all first-year bachelor students follow the same general course structure and take the same compulsory courses. 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 teaching period lasts about two months. In first-year compulsory courses, students sit centrally graded written exams at the end of each period. Starting from the second year, course grades have a participation component and other graded components, such as presentations that contribute up to 20 percent of the final courses grade.

In addition to weekly lectures, students meet in randomly assigned, group-work-based sections of 10 to 16 students. These sections are the peer groups we analyze. Section meetings last for two hours and usually take place twice per week. To prepare, students work on their study material alone or in groups before meeting with their section peers to discuss the material. These discussions are monitored and directed by an instructor, who can be a professor, lecturer, graduate, or undergraduate student.

[5] A similar description of the institutional details is provided in Elsner, Isphording and Zölitz (2018), as well as in Feld and Zölitz (2017).

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. Students or instructors do not interfere with this process. Since 2011, the random assignment has been stratified by nationality.[6] We exclude a few cases from the analysis where course coordinators did not comply with the standard allocation procedure.

The assignment of students to sections is binding. Switching from the assigned section is allowed only 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 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

Panel A of Table 1 shows basic descriptive statistics of our data. We observe 3,976 students in six subsequent cohorts. Thirty-six percent of all students are female, and students are on average 19 years old. Only 30 percent of students are Dutch, while 51 percent are German. We observe these students in 144 courses, where they are assigned to 2,522 different sections (Panel D).[7] On average, a section comprises fourteen students, and we observe students in 14.3 sections on average during their compulsory stage, 6.2 sections in the first year, and 8.1 sections in the second year. Across all

[6] In about 5 percent of the sample, 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.

[7] 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.

sections, a student meets on average 63 distinct peers during the first year and 127 distinct peers across all compulsory courses.

Educational Outcome Variables – Students Grades and Graduation: At the end of each teaching period, students sit an exam which results in a final course grade. In the first-year courses that we study, grades are exclusively based on centrally graded final exam performance. 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 the estimation sample. We further generate an indicator for graduation. In the analysis of graduation probabilities, we restrict our sample to students who could have graduated by the end of our available observation period given their enrollment year. The data does not contain high school GPA or other measures of pre-university achievement. We compute student GPA based on all grades received before the section assignment takes place. Student GPA is therefore missing for the first teaching period.

Survey Outcomes Variables – Student Course Evaluations: To compute indicators of subjectively perceived peer quality and study effort, we use students' individual-level responses to course evaluation surveys that take place at the end of each course. Course evaluation surveys are sent at the end of each course, but before students take the final exam. Responses to these are voluntary, and the response rates are about 35 percent. From the course evaluation survey, we use two variables: 1) self-reported study hours per week, excluding contact hours; and 2) a quality-of-peer interaction index as the average of the standardized values of the three evaluation items "My tutorial group has functioned well," "My fellow-students helped me to better understand the subject matter," and "The learning materials stimulated discussion with my fellow students." We standardize the quality-of-peer interaction index to have a mean of zero and a standard deviation of one over the estimation sample. Study hours keep their natural unit.

3.3 Randomization Check

The random assignment of students to sections alleviates concerns about the selection of students to peers. To check the actual randomness of peer assignment, we regress student pre-treatment characteristics (previous GPA, age, gender, and the rank of the student ID) on section dummies. This test is 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 for this property, we collect *p*-values of *F*-tests of joint significance from course-level regressions of student observables on section dummies. Under the null hypothesis of random assignment, these *p*-values are expected to be uniformly distributed, and shares of section dummies significantly explaining pre-treatment characteristics should resemble the specific confidence level (e.g., 5 percent of significant dummies at the 5 percent level).

Table 2 provides summary statistics for these balancing tests and shows that the actual rejection rates are close to the rates expected under random assignment. Moreover, *p*-values are uniformly distributed and significant at the 5 percent level in approximately 5 percent of the cases, and at the 1 percent level in approximately 1 percent of the cases. Figure A1 in the appendix shows histograms of *p*-values for the joined *F*-tests of group dummies for six student characteristics. The figure confirms that section dummies are unrelated to gender, age, GPA, and ID rank, and that the assignment is stratified based on nationality. As expected by the stratification, we observe *less* variation than expected under random assignment for the stratified variables Dutch and German.

4. Estimation of Peer Value-Added

To obtain a separate PVA estimate for each student, our empirical strategy needs to separate peer contributions from common shocks, i.e., instructor and classroom influences, as well as to isolate person-specific peer contributions from one another. Intuitively, we achieve this goal by comparing the average performance of those students who met peer j with those who did not. In the following, we explain the dyadic data structure and the three core estimation steps, which are inspired by the methods to estimate teacher value-added developed by Chetty et al. (2014a), Kane and Staiger (2008), Jacob and Lefgren (2008), Jacob, Lefgren, and Sims (2010), and Rivkin, Hanushek and Kain (2005).

Dyadic Data Structure

To implement our approach, we first reshape the original student-section-level data into a dyadic student-peer dataset. Table A1 provides an illustration of this procedure. We start from the student-course-section ($i \times c \times s$) level data, which we reshape into dyadic data at the student-peer-course-section level ($i \times j \times c \times s$) of student $\times$ peer interactions. The subscript j denotes all peers assigned to a section, and a student is not assigned to himself or herself, $i, j \in s, i \neq j$. Each student-peer pair that meets in a given section constitutes a single observation. The reshaping of the data increases the number of observations to 383,040 student-peer interactions.

We assume that all section peers are of equal relevance:

Assumption (1) *Relevant peers:* For each student i, each peer j with $i, j \in s$ and $i \neq j$ is an equally relevant peer.

We thus capture a reduced form effect of the presence of a peer on outcomes of fellow students in the same section. By maintaining this assumption, we abstain from modeling any further endogenous peer selection; for example, the endogenous formation of friendships within sections.

Step 1: Predicting Residuals

We predict residuals from a regression of grades of student i on course-times-year fixed effects. These fixed effects account for the level at which the randomization takes place. This step isolates the variation in performance that can plausibly be attributed to the changing peer composition. To increase the efficiency of our estimates, we additionally condition on student pre-assignment observables. We estimate the regression:

$$y_{ijcs} = X_i\beta + \gamma_c + \varepsilon_{ijcs} \tag{4}$$

where y_{ijcs} is the outcome (e.g., grade) of student i in section s of course c who interacts with peer j. Course fixed effects γ_c account for the level at which the randomization takes place. The vector X_i denotes pre-determined student characteristics: gender, nationality, and past GPA. From equation (4), we obtain residual student performance as:

$$y^*_{ijcs} = \widehat{\varepsilon_{ijcs}} = y_{ijcs} - \hat{\gamma}_c - X_i\,\hat{\beta} \tag{5}$$

The residual $y^*_{ijcs} = \widehat{\varepsilon_{ijcs}}$ describes a student's performance after conditioning on individual observable characteristics and demeaning at the level at which the randomization takes place. It contains peer contributions to own performance and student-related factors not captured by X_i, as well as idiosyncratic performance fluctuations at the student-section level. These residual performance measures are the building blocks of peer value-added in the next step. Note that from here onwards we drop the now obsolete subscript c after residualizing to within-course variation.

Step 2: Aggregating to Peer Value-Added

In the next step, we isolate a peer's contribution by averaging the residual performance y^*_{ijs} across all student-peer interactions with a specific peer j. To show why the average residual performance of students assigned to peer j identifies this peer's value-added PVA_j as described in Section 2, we adapt the educational production function framework outlined in equations (1) to (3) to the dyadic data and institutional setting of the Dutch business school:

$$y^*_{ijs} = PVA_j + \mu_s + \alpha_i + \epsilon_{ijs} \tag{6}$$

Here, y^*_{ijs} is the residual performance of student *i* meeting peer *j* in section *s*. This outcome is modeled as a function of peer *j*'s value-added PVA_j, remaining unobserved section-level factors μ_s, unobserved student-level factors α_i, and idiosyncratic performance fluctuations ϵ_{ijs}. Equation (6) implies that PVA_j is perfectly substitutable for the remaining inputs μ_s and α_i. To isolate PVA_j

from the remaining inputs, we assume that peers of a specific peer group affect student performance in an additive[8] way:

Assumption (2) *Functional form:* Peers *j* of a specific peer group *s* affect student *i'*s performance through an additive function of person-specific peer spillovers, namely $\sum_{j \in s, i \neq j} PVA_j$.

Averaging the residual performance y^*_{ijs} over all N_j student-peer dyads with students $i \neq j$ across the different peer groups *s* that peer *j* is observed in leads to:

$$\widehat{\mathrm{PVA}}_j = \frac{1}{N_j} \sum_{j \in s, i \neq j} y^*_{ijs} = PVA_j + \vartheta_j \tag{7}$$

$$with\ \vartheta_j = \overline{\alpha_\iota} + \overline{\mu_s} + \overline{\epsilon_{\iota J s}}$$

$\widehat{\mathrm{PVA}}_j$ measures the expected person-specific contribution of peer *j* to others' performance. The component ϑ_j contains the average student effects $\overline{\alpha_\iota}$ and the average idiosyncratic performance fluctuations $\overline{\epsilon_{\iota s}}$ of the students *i* that met peer *j*. Furthermore, ϑ_j includes average effects of the remaining peer environment $\overline{\mu_s}$. Under random assignment of students to peers, three conditions hold in expectation. First, student *i's* fixed component and peer *j*'s spillover will be orthogonal to each other: $E\left[\overline{\alpha_\iota}\, PVA_j\right] = 0$. Second, peer *j'*s spillover is unrelated to the spillovers of other classroom factors, such as teacher quality assigned to a section: E[$\overline{\mu_s}\, PVA_j$]$= 0$. Third, idiosyncratic performance fluctuations of student *i* are orthogonal to peer *j*'s

[8] Additivity is an assumption that our approach shares with linear-in-means and non-linear peer effects models commonly used in the literature.

contribution: $E[\epsilon_{ijs}\, PVA_j]= 0$. Thus, within all student-peer pairs including a given peer j, the person-specific spillover PVA_j is independent from student effects and effects of the peer environment. Therefore, averaging over all interactions with peer j will result in $\widehat{\mathrm{PVA}}_j$ as an unbiased estimate for PVA_j, the average peer-specific value-added.

Step 3: Shrinking Estimates

Our raw estimates of $\widehat{\mathrm{PVA}}_j$ measure an individual peer's expected contribution to other students' performance. We observe students in about six sections of 10 to 15 students in their compulsory stage. An individual peer meets about 63 distinct students during the first year. By chance, some peers are assigned to a group of better students, or to better teachers. Accordingly, ϑ_j picks up random fluctuations in student and peer environment quality. These fluctuations induce noise in $\widehat{\mathrm{PVA}}_j$ and might attenuate the coefficients whenever $\widehat{\mathrm{PVA}}_j$ is used as an explanatory variable.

To increase the predictive power of the PVA estimates, the third and final step is to multiply the raw $\widehat{\mathrm{PVA}}$ estimates with an estimate of their reliability. Using empirical Bayes estimates of value-added as independent variables mitigates the attenuation bias that would result from using standard OLS estimates (Jacob and Lefgren 2008; Koedel et al. 2015).

The Bayes estimator shrinks $\widehat{\mathrm{PVA}}_j$ towards zero, the prior of average PVA:

$$\widehat{\mathrm{PVA}}_{j,shrunk} \;=\; \widehat{PVA}_j\,\frac{\widehat{\sigma^2}}{\widehat{\sigma^2}+\widehat{\sigma_j^2}} \qquad (8)$$

In equation (8), $\widehat{\sigma}^2$ is the sample variance $\widehat{\sigma^2} = \frac{1}{n}\sum_j \widehat{\mathrm{PVA}}_j^2$ across all observed peers (note that the average $\widehat{\mathrm{PVA}}_j$ 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 if that peer's spillover is less consistent over time.

The resulting final shrunk estimates $\widehat{PVA}_{j,shrunk}$ defined in equation (8) are the measures that we use throughout the remainder of this paper whenever used as an independent variable on the right-hand side of a regression. When using peer value-added estimates as a dependent variable, we use the raw $\widehat{PVA}_j$ as defined in equation (7). Figure A2 in the Appendix shows a strong correlation between raw and shrunken estimates of peer value-added and an alternative value-added measure based on person-specific peer fixed effects.

4.1 Peer Value-Added and Common Shocks

Spillovers measured by a peer's PVA are robust against the problem of common shocks. As we observe students in different and 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 in which a given peer is observed. We can therefore separate the influence of peers from these idiosyncratic influences which are part of the groups-specific environment μ_s and which do not systematically vary with PVA. The intuition behind this is that, because many different interactions in many different classrooms contribute to PVA estimates, we would not expect common shocks to systematically occur in all those groups where a specific peer is present.

4.2 Peer Value-Added and the Reflection Problem

A common concern in the estimation of peer effects is the so-called reflection problem: When testing for contemporaneous effects of peers 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 pre-treatment leave-out-mean performance.

The reflection problem does not apply to the estimation of PVA. *At no stage* does the PVA approach rely on regressing student outcomes on peer outcomes. Instead, we estimate a reduced-form effect of the presence of a peer. This implies that 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.[9]

5. Results

In the following, we present our estimates of PVA. We first describe variation in PVA in our sample in Section 5.1. Here, we examine how much PVA differs between students and whether differences in PVA can be explained by observable student characteristics. We then examine pairwise correlations between PVA for different outcomes, such as contemporary grades, graduation probabilities, and perceived peer quality.

In Section 5.2, we validate our estimates as meaningful measures of peer spillovers. We show correlations of PVA measures across subjects and time and establish that PVA is predictive for out-of-sample social interactions among newly assigned peers. In Section 5.3, we provide evidence that PVA acts as a substitute for low teacher value-added. Section 5.4 shows that PVA is a malleable factor affected by past social interaction. Finally, in Section 5.5 we examine how PVA interacts with characteristics of specific student-peer matches.

[9] In Appendix 2 we empirically show that our results are not an artefact of simultaneity. We estimate PVA based on a sample where individuals enter the estimation *either* as receiving student *or* as affecting peer, but never as both. The resulting PVA remains a significant predictor in an out-of-sample prediction exercise.

5.1 Variation in Peer Value-Added

We construct PVA for several objective and self-reported outcomes. Objective outcomes are grades in the first study year, average grades in the second and third years, and graduation.[10] Self-reported measures are based on individual-level answers to teaching evaluations, and include self-reported study hours and assessments of the quality of peer-to-peer interactions.

Variation in peer value-added. Figure 1 shows the distribution of PVA for contemporary grades. The average PVA is zero by construction: We interpret PVA as the spillover of an individual peer compared to the average peer. PVA follows a fairly symmetric distribution. The two vertical lines mark a one-standard-deviation difference in PVA, which is equal to 0.036. This implies that meeting a peer with a one-standard-deviation higher PVA increases one's current grades by 3.6 percent of a standard deviation. We do not observe any bunching at the top or bottom of the distribution; rather, it appears that the majority of peers have only a small impact on others' performance. In fact, 87.5 percent of peers have a PVA that is smaller than 5 percent of a standard deviation in performance. This fact that a small number of students generates substantial peer effects is not well captured by standard peer effects models.[11]

Table 3 summarizes the variation in PVA for different student outcomes. Column (1) shows the number of students for whom we estimated PVA. Column (2) shows the across-peer standard variation in PVA. Columns (3) to (7) list various percentiles of the PVA distribution. Going beyond

[10] In general, we distinguish between two periods of the curriculum: the first study year, comprising compulsory courses only, and the subsequent second and third years, comprising both compulsory and elective courses.

[11] Hoxby and Weingarth (2005) represent a notable exception in the peer effects literature that studies how observable characteristics affect performance. The authors 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 of any moment of the distribution of observable peer characteristics.

spillovers in contemporaneous performance, meeting a peer with a one-standard-deviation higher PVA in second- and third-year grades increases these grades by 4.6 percent of a standard deviation. In the long run, meeting a peer with a one-standard-deviation higher PVA in graduation rates increases one's graduation probability by 3.5 percentage points, from a mean of 63 percent.

We also observe significant variation in PVA in self-reported outcomes based on student course evaluations. One standard deviation in PVA in self-reported study effort is equal to 0.83 study hours, or 50 minutes per week. PVA in standardized perceived peer quality is also significant, with a standard deviation of 13.5 percent. Taken together, Table 3 shows that peers have a significant impact on both objective and subjective educational outcomes. The variation in PVA is economically significant and shows that peers have a substantial impact on students' study success.

Finally, we report *p*-values of tests of joined significance of peer fixed effects in Column (8). For this purpose, we regress outcomes on peer *j* fixed effects, course fixed effects, and student background characteristics using the dyadic data. The joined significance tests reject the null hypothesis of irrelevance of peers for all objective and self-reported outcomes. Peers thus play a significant role in determining outcomes.

Peer capital. On average, a student meets about 63 distinct peers through the random section assignment during the first study year. These peers differ in their PVA. While some of them will be valuable peers and inspire higher performance, others will harm student performance compared to the average peer. We can aggregate the contributions of these multiple peers to assess the joint importance of all peers met in the first year. To quantify how much all first year peers affect a student's performance, we compute the average PVA of the pool of peers met through the random assignment. Inspired by the concept of social capital (Putnam 1995), we label this concept as *peer* *capital*. Peer capital is a student-level measure of the overall peer quality that an individual student has experienced.

The upper panel in Figure 2 shows the distribution of peer capital, with vertical lines marking the quartiles of the distribution. The lower panel of Figure 2 compares the average end-of-first-year GPA by peer capital quartile. The "lucky" quarter of students who met valuable peers from the highest quartile of the peer capital distribution on average has a 15 percent of a standard deviation higher achievement compared to the "unlucky" quarter of students with peer capital in the lowest quartile. These results highlight the importance of having good peers. Students' peer capital—the overall value-added of all peers—is a valuable asset for educational success.

Correlates of peer value-added. We next analyze which observable peer characteristics relate to a higher PVA. While we argue in Section 2 that PVA captures role model behavior, skill complementarities, or differences in socio-emotional skills, we cannot directly observe these in our data. However, we observe several peer demographic variables, namely gender, age, nationality, and the pre-assignment GPA based on all previously taken courses. Table 4 shows how these covariates relate to PVA. PVA in grades is only weakly correlated with student GPA and not correlated with gender, age, or nationality (Panel A). The *F*-Test in column (5) does not support joint significance of peer observable characteristics in predicting PVA in grades. Spillovers measured by PVA do not appear to be well captured by observable characteristics commonly used in the peer effects literature.

Observable characteristics are also poor predictors of PVA in other outcomes. Panel B shows that observables are unrelated to PVA in subsequent second- and third-year grades, graduation, study hours, and subjective peer quality. Again, *F*-tests for joint significance do not support the joint significance of student characteristics in all cases, except for the subjective peer quality index.

The weak correlation between PVA and pre-determined GPA – our best measure for student ability – deserves additional discussion. A one-standard-deviation higher pre-assignment GPA is related to 0.16 percent of a standard deviation higher PVA in grades, reflecting a small and economically insignificant relationship. To investigate whether the linear relationship between GPA and PVA is a good approximation of the functional, we provide a bin scatter in Figure 3. Figure 3 does not hint at any non-linear relationship. Apparently, it is not the high performers who bring out high performance in others, as implicitly assumed in many models of ability peer effects. However, as our data only comprises a handful of observable student characteristics and a coarse approximation of student ability, we cannot rule out that PVA is systematically related to other characteristics that are unobservable in our data.

Correlation of peer value-added across different outcomes. In the following, we examine whether PVA in different student outcomes is correlated. For example, we explore whether peers who raise performance also increase the graduation chances of fellow students. We also examine whether the presence of peers who raise performance affects how fellow students perceive the peer-to-peer interactions.

Table 5 shows pairwise correlations between PVA for different outcomes. All PVA measures for objective educational outcomes are positively correlated. Students who increase peer performance in the first year also have a positive influence on later performance on average, measured by second- and third-year grades. A peer who raises a student's first-year grades by 10 percent of a standard deviation also raises that student's later second- and third-year grades by 3.4

percent of a standard deviation on average. This correlation between contemporaneous and future effects suggests that the contemporaneous impact of first-year peers persists into later periods. This result is to some degree informative on the underlying mechanisms captured by PVA: a persistent effect suggests that PVA does not act solely through mechanisms that are dependent on the peer's presence. Part of the peer effect thus represents a persistent human capital effect that stays with the student even after the peer is no longer present. The bin scatters in Figure 4 show that we find no evidence of any non-linear relationships between our PVA measures.

We also find a positive correlation between PVA in first-year grades and graduation probabilities. A 10-percent increase in PVA in first-year grades is related to a 2.7 percent increase in graduation chances, reflecting a small yet statistically significant correlation. Similarly, peers who raise subsequent second- and third-year performance on average also raise graduation probabilities.

Correlations between objective and subjective outcomes are smaller than between objective outcomes. Peers who raise student performance by 10 percent of a standard deviation raise students' perceptions about the quality of peer-to-peer interactions by 0.4 percent of a standard deviation on average. While the exact content of PVA is unobservable to the econometrician, students notice valuable peers in the classroom and evaluate peer-to-peer interaction more positively in the student course evaluations.

PVA in contemporary performance is *negatively* correlated to PVA in study hours. Those peers who raise the performance of their fellow students seem to reduce average effort provision. While self-reported study hours are a noisy measure of how much students study, this negative relationship points to a potential substitution effect of peer-supported and own learning. Students who meet peers with higher PVA substitute own study efforts with their peers' ability to teach them the course material, or become more efficient in their self-study.

5.2 Validating Peer Value-Added

After describing the variation in PVA, we now investigate its stability over time and across subjects. Furthermore, we test the predictive power of PVA in out-of-sample predictions of performance spillovers among newly assigned peers. There are two reasons the stability and predictive power is of interest. First, PVA might comprise a context-independent and a transitory component. Correlations across context and time tell us about the size of the context-independent component in PVA relative to the transitory component. Second, as PVA is based on a finite number of student-peer matches, it may still contain noise. The correlation of PVA across subjects and periods is informative about the amount of signal. If PVA contains no systematic information and only captures random fluctuations in performance, it should not be correlated across subjects and time periods.

Stability of peer value-added. PVA estimated separately on first- vs. second- and third-year student-peer pairs is significantly and positively correlated (Figure 5, upper panel). Regressing PVA based on first-year section assignments on PVA based on second-/third-year assignments yields a coefficient of $\hat{\beta} = 0.11$. This correlation is smaller than what is typically found in the teacher value-added literature, and hints at a substantial transitory component in PVA. This relatively low correlation could be rationalized by students taking courses in a variety of subjects, learning new content, and interacting with new peer groups in every new course that they take. Notably, the beginning of university studies is a formative period for many students, during which they could also systematically alter how much they affect the learning of other students. For teachers, by contrast, many aspects of the environment do not dramatically change from one year to the next. They often teach the same subjects and content in the same room, with the same colleagues.

We further assess the stability of PVA across courses of different subjects. For this purpose, we separately estimate PVA based on math-intensive and non-math-intensive courses in the first year, and test how strongly these are correlated. The lower panel of Figure 5 shows the relationship between PVA in mathematical versus non-mathematical courses. Regressing PVA in non-math-intensive subjects on PVA in math-intensive subjects yields a coefficient of $\hat{\beta} = 0.21$, which is significant at the 1 percent level and stronger than the year-to-year correlation. PVA in different subjects is related, albeit imperfectly. On average, a student who raises peers' performance in math-intensive subjects also increases fellow students' performance in non-math-intensive courses.

Predictive power for future interactions. To serve as a useful tool to detect valuable peers, PVA should be able to predict a peer's spillover in a new sample different from the one in which it was estimated. Predictive power is the best evidence for PVA picking up actual variation in peer spillovers. Further, out-of-sample predictive power is a prerequisite for deriving valid policy recommendations from peer effects estimates (Carrell, Sacerdote and West 2013). To assess the predictive power of our PVA measures, we exploit the fact that students are again randomly allocated to new peer groups in their second- and third-year courses.[12] We test whether PVA – measured on first-year courses – is predictive for outcomes when interacting with these newly

[12] Table A2 in the Appendix provides an additional randomization check for the out-of-sample prediction sample, and confirms that student characteristics are unrelated to peer value-added in this new sample.

assigned peers. For this purpose, we regress student-level standardized course grades obtained in the second and third year on PVA using student-peer dyad data. Notably, the second- and third-year student grades are obtained among newly assigned peers different from those on which PVA is based. We thus test whether PVA can predict student performance *out-of-sample*. Figure 6 visualizes the regression results of this exercise, and shows that PVA measured in the first year is a valid predictor of the performance of newly assigned students in later years. A one-standard-deviation increase in first-year PVA raises the performance of newly assigned students by 9.9 percent of a standard deviation. The size of this coefficient is comparable to the degree of over-time and across-subject reliability of PVA found in the previous section.

5.3 Interaction with Teacher Value-Added

Just like their students, instructors are randomly assigned to sections within courses, which allows us to investigate interaction effects between PVA and teacher value-added. This interaction of peer and teacher effect – two key determinants of educational performance – has received very little attention in the literature. For this purpose, we estimate teacher value-added analogously to the methodology used for PVA for the university instructors in our sample.[13] We base this exercise on regressions of second- and third-year grades on PVA interacted with tertiles of the teacher value-added distribution. The regressions are again based on out-of-sample predictions. Both peer and teacher value-added measures are pre-determined and based on a different sample than the estimation sample.

[13] Teacher value-added can only be computed when teachers are observed in multiple courses. Thus, teacher value-added is available for 80 percent of the validation sample. Our teacher value-added estimates are close to Feld, Salamanca and Zölitz (forthcoming), who estimate teacher value-added in the same setting following the methods of Chetty et al (2014b).

Figure 7 shows the effect of PVA interacted with tertiles of the teacher value-added distribution. The horizontal line indicates the point estimate from the overall sample. Almost the entire average effect of PVA is driven by sections taught by low value-added instructors.[14] The results point to a substitution effect between peer and teacher quality. It appears that students can rely on the supportiveness of their peers when facing teachers of poor quality. Put differently, peers matter less in classrooms where high-quality teachers are present. This interaction between PVA and teacher quality suggests that student-teacher reassignment policies could increase overall performance.

5.4 Malleability of Peer Value-Added

We now ask whether students become better in raising others' performance after being exposed to valuable peers in earlier periods. We investigate this by again taking advantage of the out-of-sample exposure to different peer groups over the course of study. If PVA is malleable, interventions targeting peers instead of students themselves could raise overall performance.

To test whether peers affect a student's own PVA, we regress person-specific PVA constructed on second- and third-year courses on peer capital, the average PVA of all peers met in the first year. Figure 8 shows that PVA changes systematically after exposure to higher value-added peers in earlier periods. A one-standard-deviation increase in students' peer capital raises their own subsequent PVA in grades by 4.5 percent of a standard deviation.

There are several potential mechanisms behind this malleability of PVA that are not observable in our data. Students might adapt to peer behavior that they observed and benefited

[14] Table A3 in the Appendix shows the underlying regression results and robustness of the results toward the inclusion of instructor fixed effects.

from in earlier periods. They may learn to ask more informative questions, and be more supportive during class, or help their peers to better study for the exam. These "learning effects" suggest that students who meet high value-added peers not only perform better but subsequently also produce larger spillovers, whereby they become better at raising other students' performance. The malleability of PVA is similar to results of Jackson and Bruegmann (2009), who show that teacher value-added increases after exposure to more effective, high value-added colleagues.

5.5 Heterogeneity by Student-Peer Match Characteristics

The impact of a peer might not be fully homogenous across social interactions. While some students may – for example – be better at helping academically-weak students, others may have a strong impact on students of the same gender. Such heterogeneity in the effect of value-added by match characteristics is not well captured by average PVA estimates. However, knowledge about such heterogeneities may become important when re-assigning peers to increase overall performance.

To understand how general the impact of PVA is, we regress student grades obtained in the second and third year on PVA and the interaction between PVA and match characteristics. As match characteristics, we consider whether peer and student have the same gender and same nationality, and whether they have similar past performance measured as less than 50 percent of a standard deviation difference in their past GPA. Table 6 shows the estimation results on how first-year-based PVA interacts with student-peer match characteristics. Although confidence intervals are largely overlapping, peer spillovers appear stronger in homogeneous student-peer matches. This result is consistent with Opper (2019), who also finds larger spillovers for same race and gender interactions. Matching characteristics in themselves have only very small and mostly insignificant positive effects. More homogeneous classes in terms of gender, nationality, or student achievement appear somewhat to foster spillovers between students.

6. Discussion

6.1 Relationship with the Teacher Value-Added Literature

How do our PVA measures relate to the results in the teacher effectiveness literature? Figure 9 compares our PVA estimates to estimates in the literature on the value-added of teachers and school principals. The variation in PVA is significantly smaller than the variation estimated for teacher or principal value-added. This implies that, in expectation, replacing a single teacher will have a larger effect on performance than replacing a single peer. In Figure 9, the median variation in teacher value-added is 0.11 of a standard deviation in student performance. Our .036 estimate of PVA is about one-third of these median estimates.

Some of our findings for PVA closely resemble findings in the teacher value-added literature: peer observables are poor predictors for PVA, and exposure to high value-added peers increases own subsequent value-added among newly-assigned peers. Jackson and Bruegmann (2009) document similar learning spillovers among teachers' value-added. Like teacher value-added, PVA is systematically correlated over time and context. This correlation is lower than correlations in the teacher value-added literature.

6.2 Relationship with Prior Work on Peer Effects

Our findings have several implications for how to interpret existing findings from the literature on peer spillovers. The large variation in PVA indicates that flexible functional forms are necessary to account for the role of "shining lights" and "bad apples" in generating peer spillovers. For example, focusing on average peer group characteristics is likely to overlook peer influences driven by a few high value-added peers.

Second, 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 necessarily capture the relevant dimensions in which students affect their peers. Estimates of peer effects in single observable dimensions might not be informative about the lack or presence of peer effects in general when the ability to raise others' performance is largely orthogonal to observables. More broadly, the absence of spillovers generated by peer observable characteristics does not necessarily imply the absence of peer effects in general. The surprisingly small magnitude or entire lack of peer spillovers found in many settings might be misleading regarding the true existence and size of peer spillovers.

Our paper complements three other studies that have acknowledged the potential role of unobservable peer characteristics without estimating peer value-added. Cooley 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 ability.[15] Closest to our paper, Arcidiacono et al. (2017) estimate player-level contributions to others' performance based on National Basketball Association data. Lacking information on individual player productivity, however, they cannot rely on direct player-peer pair observations, and selection of players into the game cannot be ruled out.

[15] Note that this approach is significantly different from our strategy. Arcidiacono et al. (2012) model peer spillovers as function of unobservable peer ability, proxied by peer fixed effects from a regression of peer j 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 this peer's ability. In fact, our results do not show a strong relationship between student ability and PVA.

The authors solve these issues by relying on a saturated structural model of player productivity, recovering estimates of player-specific spillovers.

6.3 Applications in Education Policy and Human Resources Management

The methods developed in this paper can be applied to other settings with output-based incentives or performance feedback. In many education and workplace settings, agents are awarded for individual output, but not for how much they contribute to the performance of others. When peer spillovers are substantial, worker-specific performance measures insufficiently describe one's actual contribution to team production. Incentives for 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 potentially be used to adjust worker-specific productivity measures for contributions that should be attributed to specific co-workers. Knowing who valuable co-workers are could help to allocate workers to output-maximizing teams. The approach we developed in this paper could also help to ascertain whether workers sort into firms based on co-worker value-added, and whether firms pay workers according to their co-worker value-added.

7. Conclusion

In this paper 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 to his/her fellow students' performance, which we label the peer value-added. In contrast to the existing measures of peer effects, peer value-added summarizes the total contribution to others' performance capturing both observable and unobservable characteristics that create spillovers. Therefore, the peer value-added approach does not require taking a stand about which peer observables create spillovers.

Our results show significant variation in how much peers affect university performance. The majority of peers have only a small impact on performance: out of all peers with whom students interact, only 12.5 percent affect grades by more than 5 percent of a standard deviation. However, the overall pool of peers that a student meets during the first year has a substantial impact on performance. We define the peer capital of a student as the average peer value-added of all students met throughout the first year. A student in the top quartile of the peer capital distribution has on average 15 percent of a standard deviation higher end-of-first-year GPA than a student in the bottom quartile of the peer capital distribution.

When looking at what makes a peer valuable, we find that peer value-added is largely unrelated to peers' observable characteristics. Most notably, peer value-added is only very weakly correlated with past achievement. Peer value-added is correlated across time and subjects. Furthermore, peer value-added is a valid out-of-sample predictor for spillovers among newly assigned peers in later years.

Peer value-added acts as a substitute of low teacher quality. Valuable peers affect student performance more in classes taught by less effective teachers. Peer value-added further proves to be malleable: changes in peer value-added can be partly explained by past exposure to high value-added peers. Students thus learn "how to be a good peer" from interacting with valuable peers who bring out the best in others.

The conceptual framework and estimation that we propose in this paper provide a novel perspective on how we think about peer effects. Peer value-added allows estimating peer effects without priors about which observable peer characteristics generate spillovers. In fact, 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 performance observation and exogenous variation in group composition. Peer value-added may therefore serve as a tool to re-assess the importance of peers in a variety of education and workplace settings.

REFERENCES

Angrist, J. D. (2014). The Perils of Peer Effects. *Labour Economics*, *30*, 98–108.

Arcidiacono, P., Foster, G., Goodpaster, N., 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.

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. *NBER Working Paper 17803*.

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., Hoekstra, M., and Kuka, E. (2018). The Long Run Effects of Disruptive Peers. *American Economic Review*, 108(11) 3377-3415.

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.

Cooley Fruehwirth, J. (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

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., and Zölitz, U. (2018). Achievement Rank Affects Performance and Major Choice in College. *University of Zurich Working Paper Series 300*.

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. (forthcoming) Are Professors Worth It? The Value-Added and Costs of Tutorial Instructors. *Journal of Human Resources*.

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., 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*.

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.

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.

Opper, I. M. (2019). Does Helping John Help Sue? Evidence of Spillovers in Education. *American Economic Review*, 109 (3), 1080–1115.

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.

Tincani, M. M. (2018). Heterogeneous Peer Effects in the Classroom. *Working Paper*.

Sacerdote, B. (2014). Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward? *Annual Review of Economics*, 6(1), 253–272.

Wang, L. C. (2010). Peer Effects in the Classroom: Evidence from a Natural Experiment in Malaysia. *Unpublished manuscript*.

TABLES AND FIGURES

Tables

Table 1: Descriptive Statistics

	(1) N	(2) Mean	(3) Sd	(4) Min	(5) Max
A. Student Demographic Characteristics					
Female	3,976	0.359	0.480	0	1
Dutch	3,976	0.303	0.460	0	1
German	3,976	0.514	0.500	0	1
Age	3,976	19.15	1.590	16.19	39.73
B. Student Performance					
Pre-assignment GPA	25,845	6.681	1.464	1	10
Course grade in year 1	23,343	6.216	1.774	1	10
Std. average grade in years 2 & 3	19,971	6.815	1.181	1	9.73
Graduation	15,758	0.634	0.482	0	1
C. Students' Course Evaluations					
Self-reported study hours	8,358	11.97	7.833	0	70
Std. subjective peer quality index	7,889	-0.004	1.001	-3.70	1.77

D. Summary Statistics of Dyadic Student-Peer Interactions	N
Number of individuals	3,976
Number of sections (peer groups)	2,522
Number of course-year observations	144
Total number of dyadic student-peer interactions	383,040
Average number of dyadic interactions per group	151.88
Average number of students per section	13.98
Number of unique peers students meet in the first year	62.80
Number of unique peers students meet during their studies	126.88
Average number of student-peer interactions conditional on meeting once	1.21

Note: This table summarizes descriptive statistics of the estimation sample. 'Sd' refers to the standard deviation of the respective variable. Panel A reports individual student characteristics. Panels B and C report outcomes at the student-course level. Smaller numbers of observations in Panel B are due to student dropout. Smaller numbers of observations in Panel C are due to non-response to teaching evaluations. Self-study hours are based on the question, "How many hours per week on the average (excluding contact hours) did you spend on self-study?" The subjective peer quality index is based on the evaluation items "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," and "The learning materials stimulated discussion with my fellow students." Panel D shows the descriptive statistics of the estimation sample before and after reshaping the data to student-peer dyads.

Table 2: Test for Random Assignment of Students to Sections

	(1)			(2)			(3)
	Number significant at the:			Percent significant at the:			Number of total
Dependent variable:	5%	1%	0.1%	5%	1%	0.1%	tests performed
Female	2	0	0	1.5%	0.0%	0.0%	136
GPA	8	0	0	5.6%	0.0%	0.0%	144
Age	4	1	0	2.8%	0.7%	0.0%	144
ID rank	5	3	0	3.5%	2.1%	0.0%	144
Dutch	0	0	0	0.0%	0.0%	0.0%	138
German	3	1	0	2.2%	0.7%	0.0%	137

Note: This table is based on separate OLS regressions with gender, GPA, age, and ID rank as dependent variables. The explanatory variables are a set of section dummies and dummies for parallel courses taken simultaneously. Columns (1) and (2) show in how many regressions the *F*-Test 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 stratification of assignment by nationality. Figure A1 shows histograms of *p*-values of each performed test from the underlying regressions.

Table 3: Summary Statistics of Peer Value-Added Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Percentiles					*p*-value *F*-Test of joined sig. of peer FE
Peer Value-Added Estimate	N	SD	p1	p5	p50	p95	p99	
Performance indicators								
Std. course grade 1st year	3,976	0.0356	-0.1054	-0.0557	0.0003	0.0566	0.1076	0.0120
Std. average course grade 2nd & 3rd years	3,914	0.0456	-0.1351	-0.0623	0.0012	0.0819	0.1371	<0.001
Graduation	3,224	0.0350	-0.1120	-0.0495	0.0002	0.0530	0.0973	<0.001
Self-reported measures								
Self-study hours	3,939	0.8293	-2.5072	-1.5628	-0.0323	1.0196	2.0612	<0.001
Subjective peer quality	3,906	0.1351	-0.3815	-0.1926	0.0015	0.2392	0.4131	<0.001

Note: This table reports summary statistics of value-added estimates at the peer level. Underlying grades are standardized to mean zero and unit variance. Self-study hours are kept in their natural unit and based on the question, "How many hours per week on the average (excluding contact hours) did you spend on self-study?" Subjective peer quality is an index standardized to mean zero and unit variance based on the evaluation items "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," and "The learning materials stimulated discussion with my fellow students." The number of observations differ due to missing values of outcomes. In column (8) we report *p*-values from joint significance tests performed after regressions of outcomes on peer fixed effects based on dyad data.

Table 4: Who is a Good Peer?

Correlation between Student Characteristics and Various Peer Value-Added Measures

Panel A	(1)	(2)	(3)	(4)	(5)
Dependent Variable: Peer Value-Added	Std. course grade	Std. course grade	Std. course grade	Std. course grade	Std. course grade
GPA	0.0016*				0.0018**
	(0.001)				(0.001)
Female		0.0027			0.0026
		(0.003)			(0.003)
German			-0.0023		-0.0046
			(0.004)		(0.004)
Dutch			-0.0044		-0.0038
			(0.004)		(0.004)
Age				0.0012	0.0015
				(0.001)	(0.001)
Observations	3,976	3,976	3,976	3,976	3,976
R-squared	0.001	0.000	0.000	0.000	0.002
p-value of *F*-Test for joint significance of student characteristics	-	-	-	-	.1837

Panel B					
Dependent Variable: Peer Value-Added	Std. course grade	Std. average 2nd and 3rd year grades	Graduation	Self-study hours	Subjective peer quality
GPA	0.0018**	-0.0006	-0.0005	0.0091	0.0095***
	(0.001)	(0.001)	(0.001)	(0.019)	(0.003)
Female	0.0026	0.0021	0.0037	0.0022	0.0002
	(0.003)	(0.004)	(0.003)	(0.060)	(0.010)
German	-0.0046	0.0023	-0.0028	-0.0115	-0.0236*
	(0.004)	(0.005)	(0.004)	(0.080)	(0.013)
Dutch	-0.0038	0.0038	0.0010	0.0294	-0.0051
	(0.004)	(0.005)	(0.004)	(0.085)	(0.014)
Age	0.0015	0.0011	0.0012	0.0106	0.0028
	(0.001)	(0.001)	(0.001)	(0.019)	(0.003)
Observations	3,976	3,944	3,536	3,973	3,975
R-squared	0.002	0.000	0.001	0.000	0.003
p-value of *F*-Test for joint significance of student characteristics	.1837	.9001	.4333	.9738	.0206

Note: This table reports OLS regression results of PVA in different outcomes on peer observable characteristics. One observation per peer. Robust standard errors are in parentheses. Significance levels indicated as * p<0.1, ** p<0.05, and *** p<0.01.

Table 5: Correlation between different Peer Value-added Measures

Pairwise Regressions of Peer Value-Added Measures	(1) Contemporaneous Course Grade	(2) Average 2nd and 3rd Year Grades	(3) Graduation	(4) Self-study Hours
Average 2nd and 3rd year grades	0.3386***			
	(0.029)			
Graduation	0.2727***	0.6708***		
	(0.041)	(0.051)		
Self-study hours	-0.0033**	0.0060***	-0.0002	
	(0.002)	(0.002)	(0.002)	
Subjective peer quality	0.0403***	0.0554***	-0.0139	0.3599*
	(0.010)	(0.013)	(0.010)	(0.212)

Note: This table reports results from pairwise regressions of PVA in different outcomes on the level of individual peers. Self-study hours are kept in their natural unit and based on the question, “How many hours per week on the average (excluding contact hours) did you spend on self-study?” Subjective peer quality is an index standardized to mean zero and unit variance based on the evaluation items “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,” and “The learning materials stimulated discussion with my fellow students.” One observation per student. Robust standard errors are in parentheses. Significance levels indicated as * p<0.1, ** p<0.05, and *** p<0.01.

Table 6: Validation of Peer Value-Added – Predictive Power of PVA for New Interactions

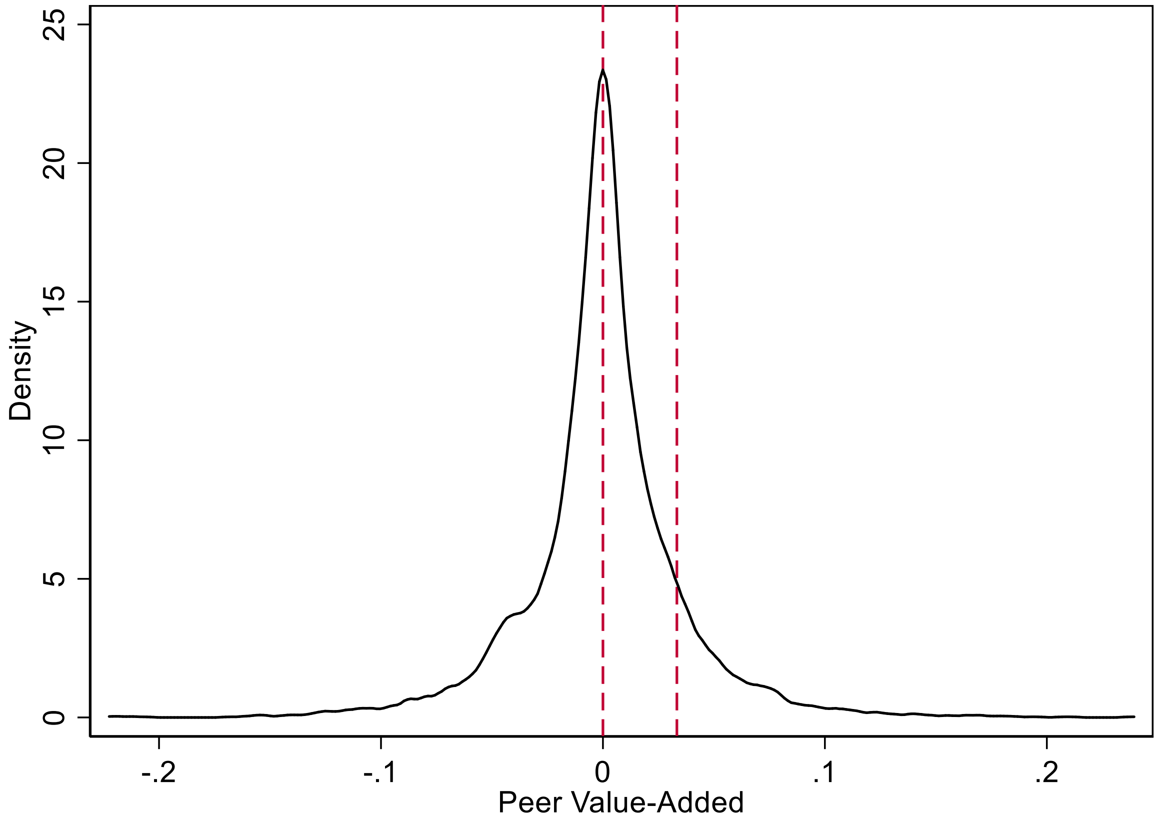
Does the Impact of Peer Value-Added Depend on the Student-Peer Match?

Dependent Variable: Std. Course Grade in 2nd/3rd courses	(1) Std. course grade	(2) Std. course grade	(3) Std. course grade	(4) Std. course grade	(5) Std. course grade
Peer Value-Added	0.0992***				
	(0.033)				
Peer Value-Added * Same gender match		0.1172**			
		(0.049)			
Peer Value-Added * Opposite gender match		0.0782			
		(0.057)			
Peer Value-Added * Same nationality match			0.1394**		
			(0.061)		
Peer Value-Added * Different nationality match			0.0692		
			(0.048)		
Peer Value-Added * Same gender & nationality match				0.1696*	
				(0.090)	
Peer Value-Added * Different gender or nationality match				0.0780*	
				(0.042)	
Peer Value-Added * GPA match					0.1332*
					(0.071)
Peer Value-Added * no GPA match					0.0865**
					(0.039)
Same gender match		0.0038			
		(0.003)			
Same nationality match			0.0087**		
			(0.004)		
Same gender & same nationality match				0.0053	
				(0.004)	
GPA match					0.0031
					(0.005)
Observations	312,240	312,240	312,240	312,240	312,240
R-squared	0.461	0.461	0.461	0.461	0.461

Note: This table reports OLS regression results of students' grades in second- and third-year sections on peers' PVA based on first-year interactions. Interacted variables refer to the respective matches in observable characteristics between peer and student. One observation is one student-peer dyad in second- and third-year courses. GPA match is an indicator variable that equals one if the difference between student and peer GPA is less than 50 percent of a standard deviation. Additional controls include gender, nationality and quartic polynomials of GPA. Robust standard errors using two-way clustering at the individual and section levels are in parentheses. Significance levels indicated as * p<0.1, ** p<0.05, and *** p<0.01.

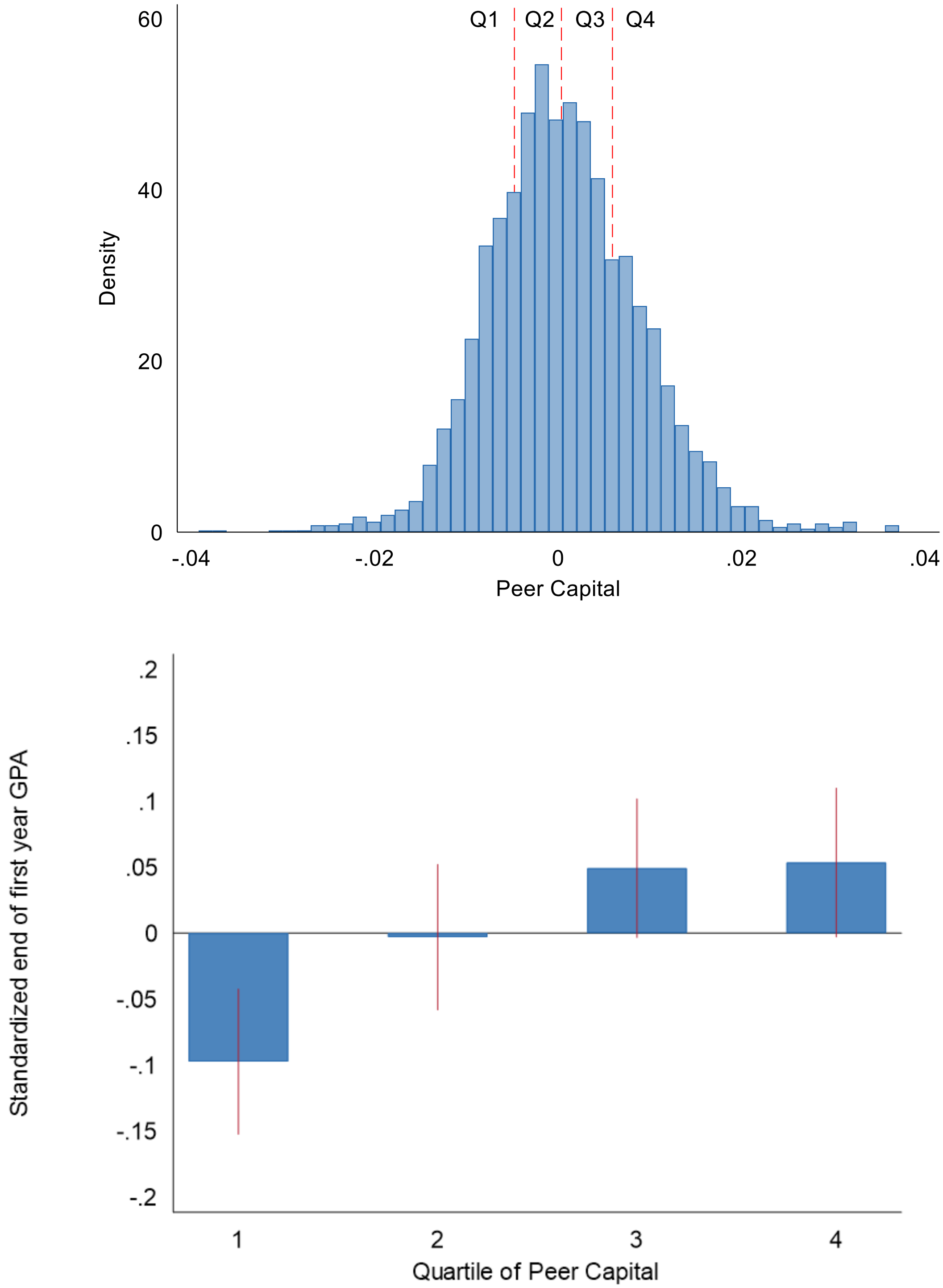
Figures

Figure 1: Distribution of Peer Value-Added



Note: This figure shows the distribution of estimated peer value-added for standardized course grades in the first year. The distance between the two vertical lines is one standard deviation in peer value-added. This is equivalent to 3.6 percent of a standard deviation in first-year grades. Out of all peers 12.5 percent affect performance by more than 5 percent of standard deviation. N=3,976.

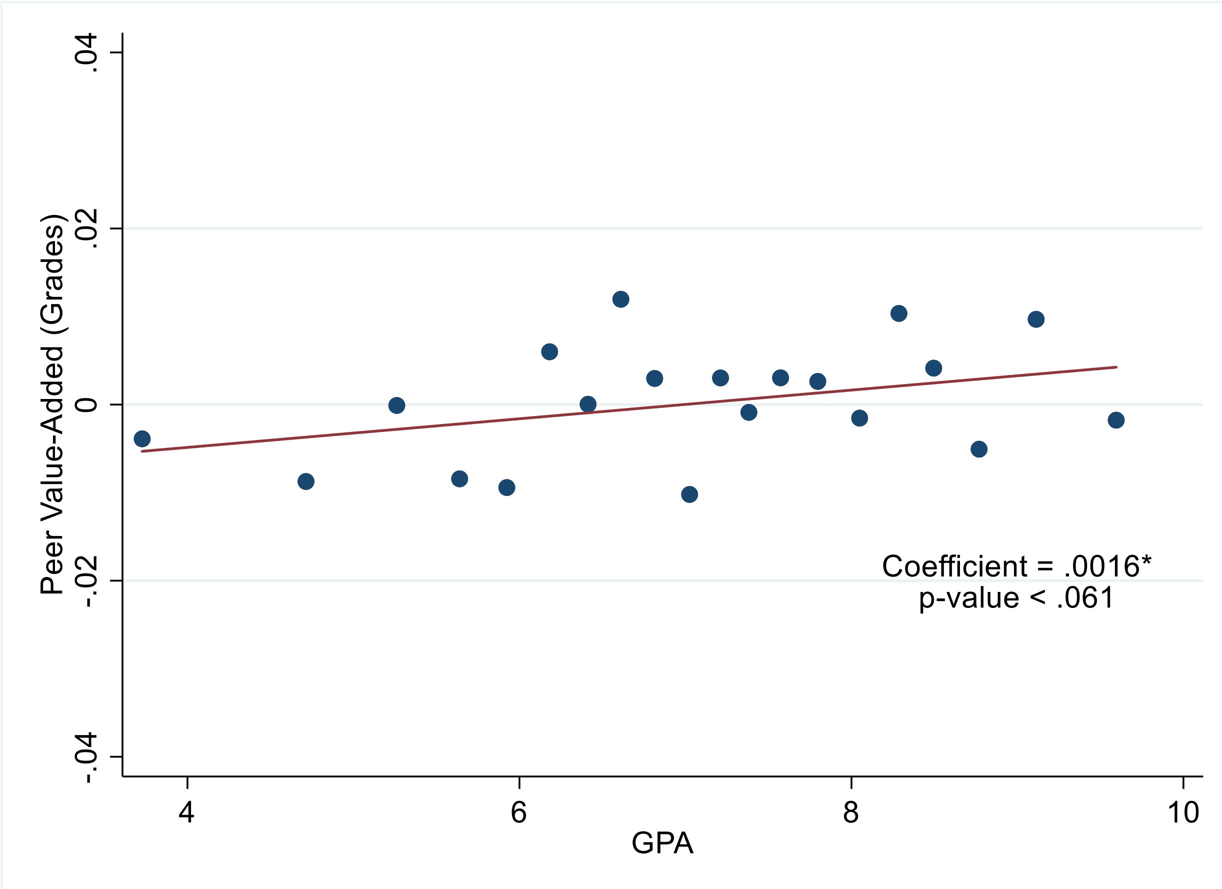
Figure 2: Peer Capital Defined as the Mean of Peer's Value-Added



Note: Panel A shows a histogram of the distribution of *peer capital*. Peer capital is defined as the average peer value-added of peers met by a specific student in the first study year. The vertical lines mark the 25th, 50th and 75th percentile of the peer capital distribution.

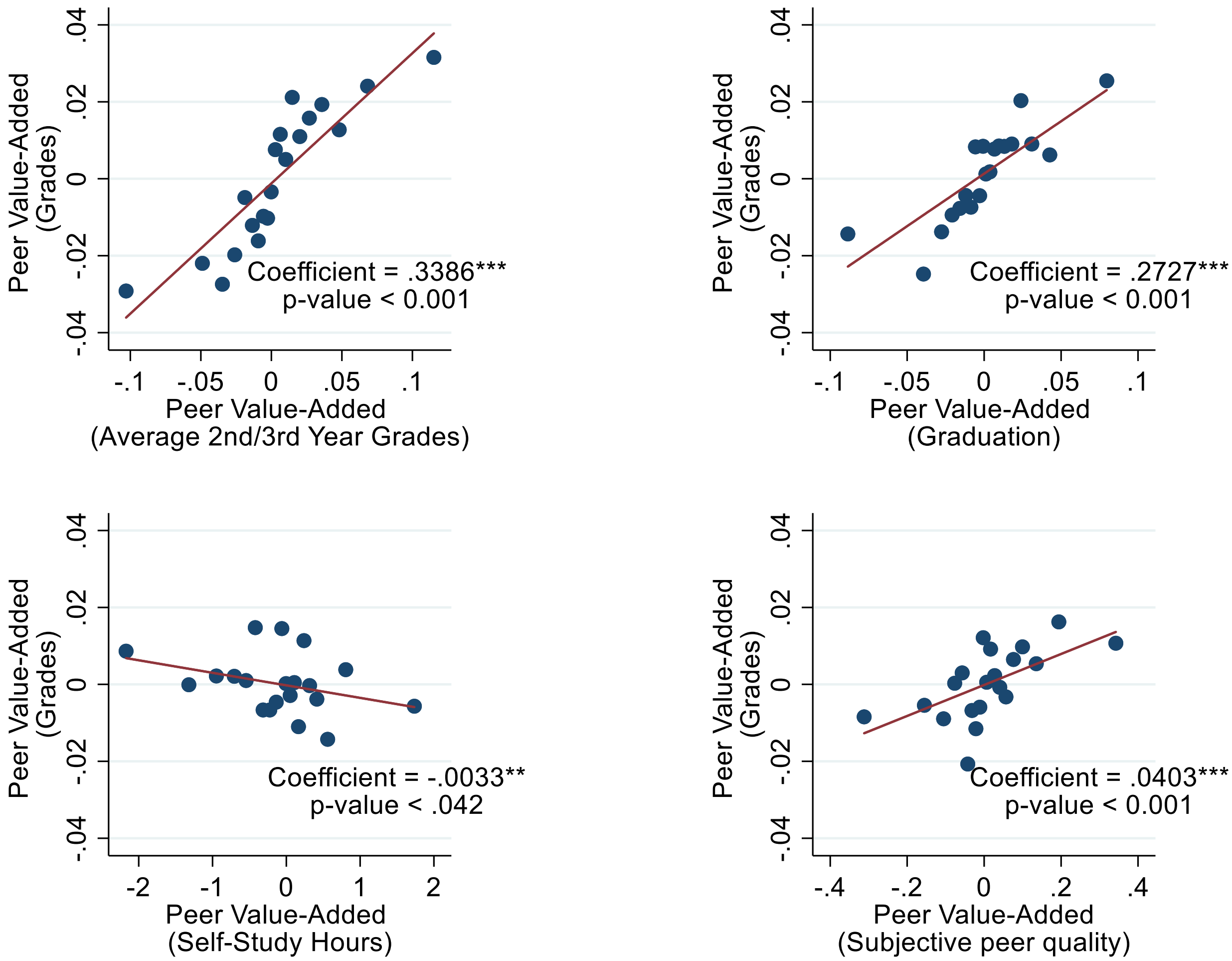
Panel B shows the standardized first-year GPA, split by student quartile of peer capital. At the end of the first year, students in the top quartile of the peer capital distribution have a 15 percent of a standard deviation higher GPA than students in the bottom quartile.

Figure 3: Correlation between Student GPA and Peer Value-Added



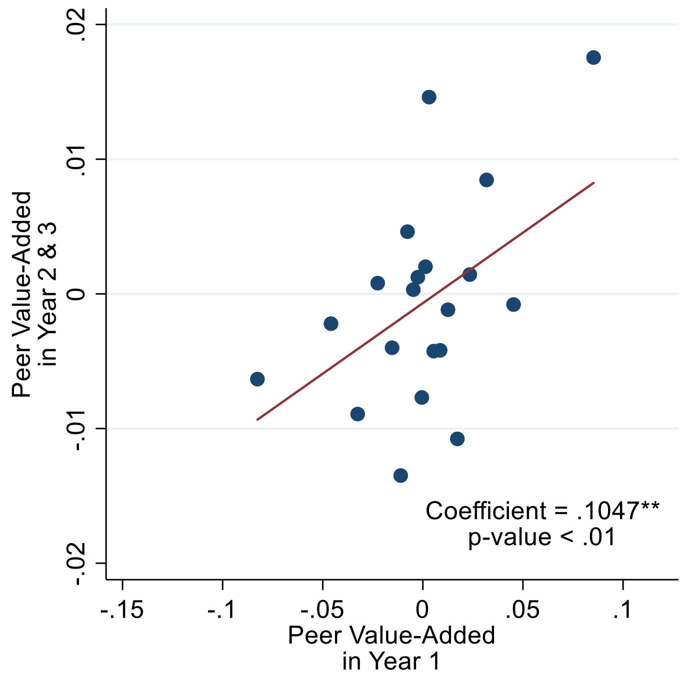
Note: This bin scatter visualizes the relationship between PVA and past GPA of a peer. The point estimate and significance level are obtained from the OLS regression reported in Table 4. N = 3,976. Significance levels indicated as * p<0.1, ** p<0.05, and *** p<0.01.

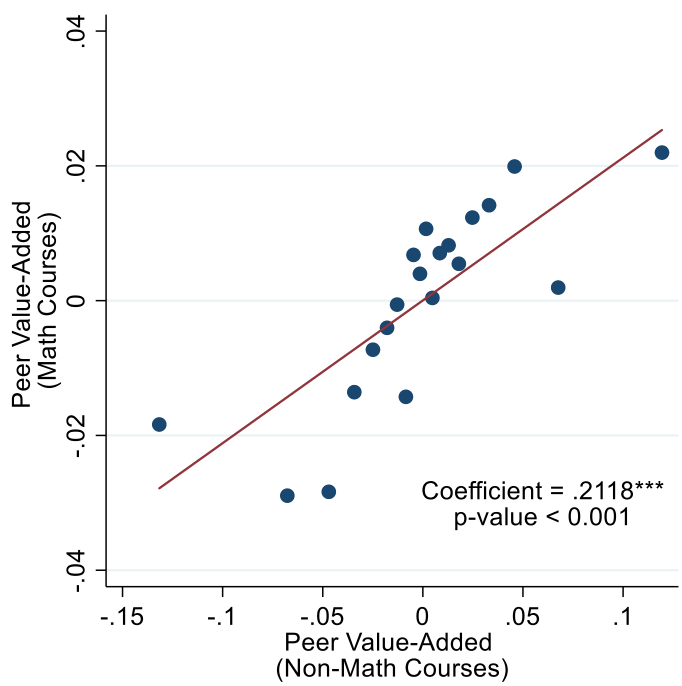
Figure 4: Correlation between Peer Value-Added in Grades and PVA in Other Outcomes



Note: These bin scatters show how PVA in grades is related to PVA in other outcomes. Point estimates and significance levels are obtained from pairwise OLS regression reported in Table 5. One peer is one observation. N_1 = 3,914; N_2 = 3,224; N_3 = 3,939; N_4 = 3,906. All *p*-values are based on robust standard errors. Significance levels indicated as * p<0.1, ** p<0.05, and *** p<0.01.

Figure 5: Stability of Peer Value-Added Measures across Time and Subjects

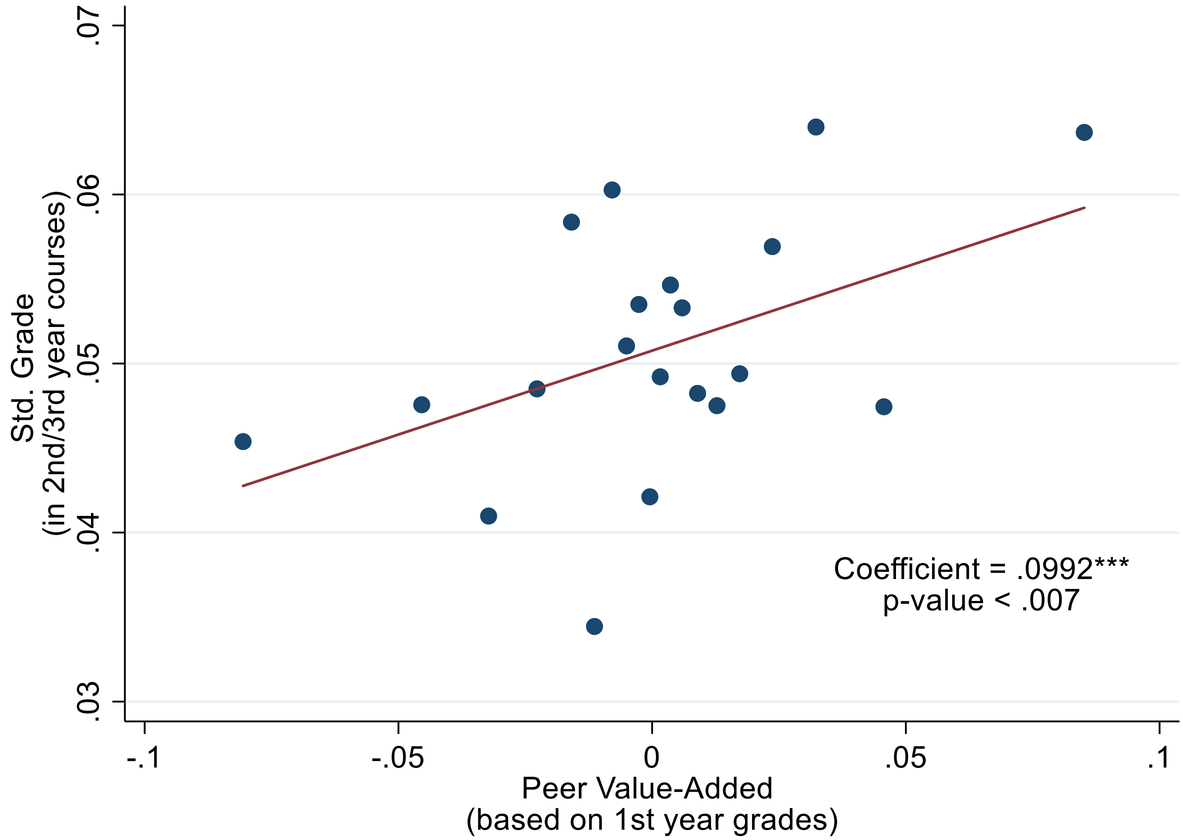




Note: The top figure shows the peer-level relationship between PVA separately estimated on student-peer dyads in the first vs the second andthird years. The bottom figure shows the peer-level relationship between separately estimated PVA in math- and non-math-intensive courses. Point estimates and significance levels are obtained from OLS regressions. N1 = 3,237; N2 = 3,935; *p*-values are based on robust standard errors. Significance levels indicated as * p<0.1, ** p<0.05, and *** p<0.01.

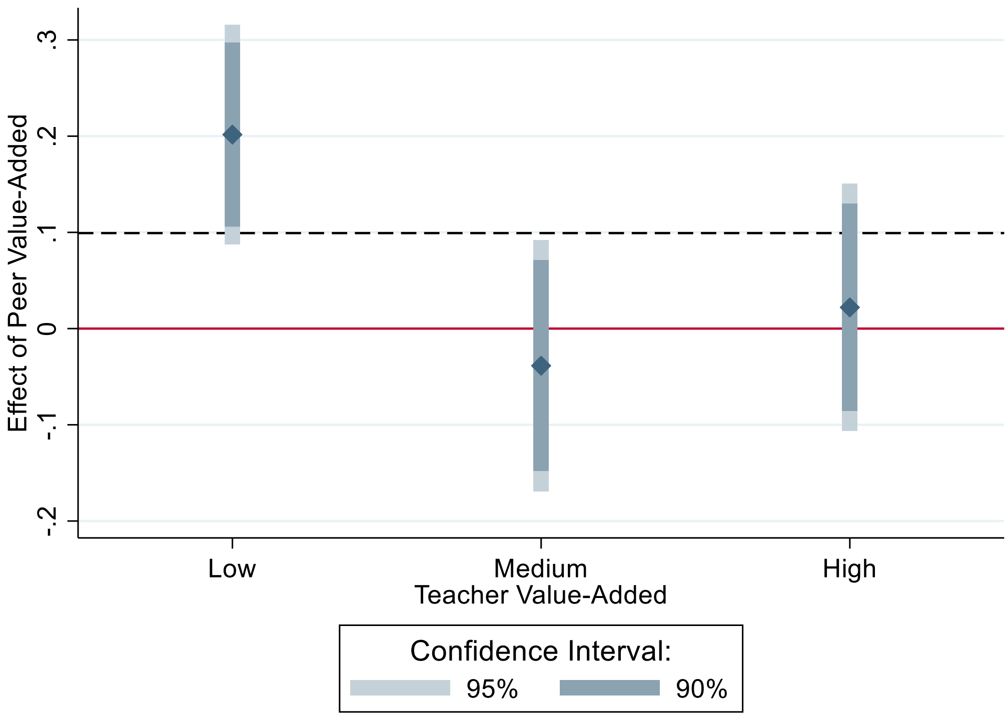
Figure 6: Out-of-Sample Prediction –

Effect of Peer Value-Added on Newly Assigned Students



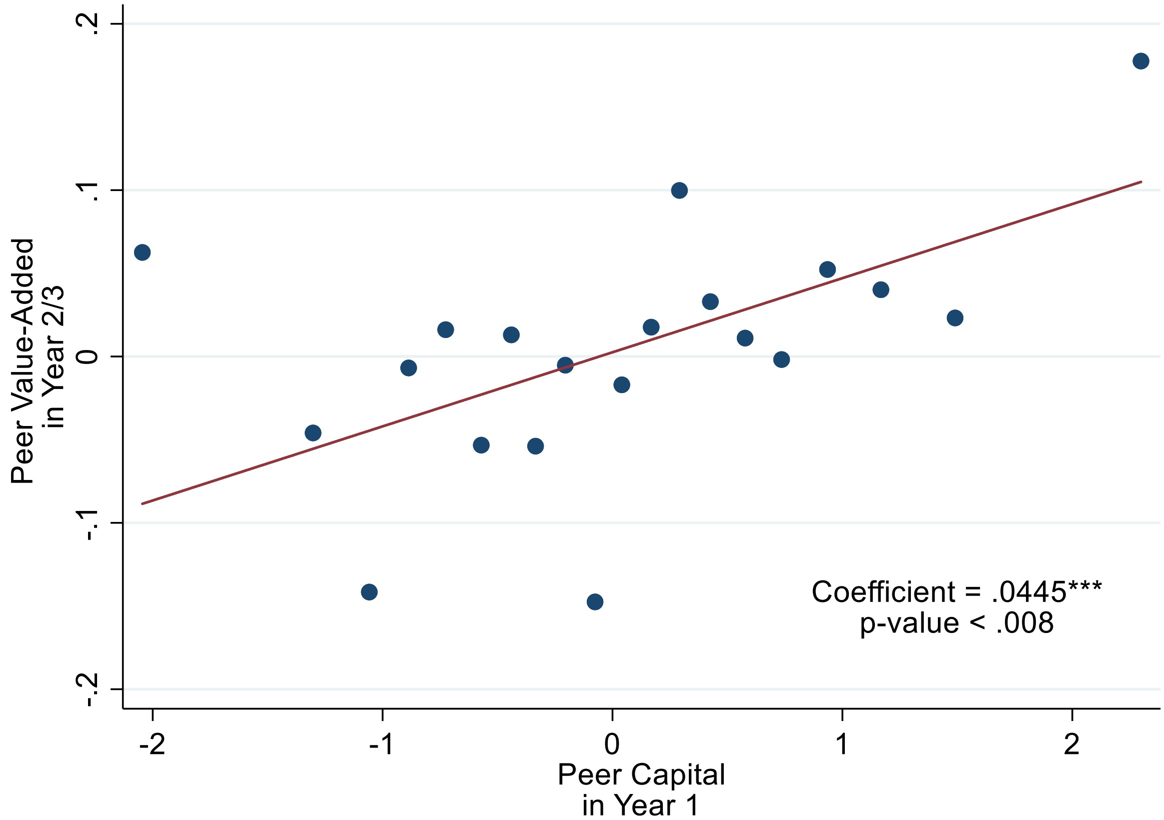
Note: This bin scatter visualizes the relationship between own grades and the peer value-added of section peers. The point estimate and significance level are obtained from the OLS regression reported in Table 6. The estimation sample consists of all student-peer interactions in second- and third-year courses. The peer value-added measure is constructed based on first-year student-peer pairs. One observation is one student-peer interaction. N=316,382. All *p*-values are based on two-way clustering at the individual and section levels. Significance levels indicated as * p<0.1, ** p<0.05, and *** p<0.01.

Figure 7: Interaction between Teacher and Peer Value-Added



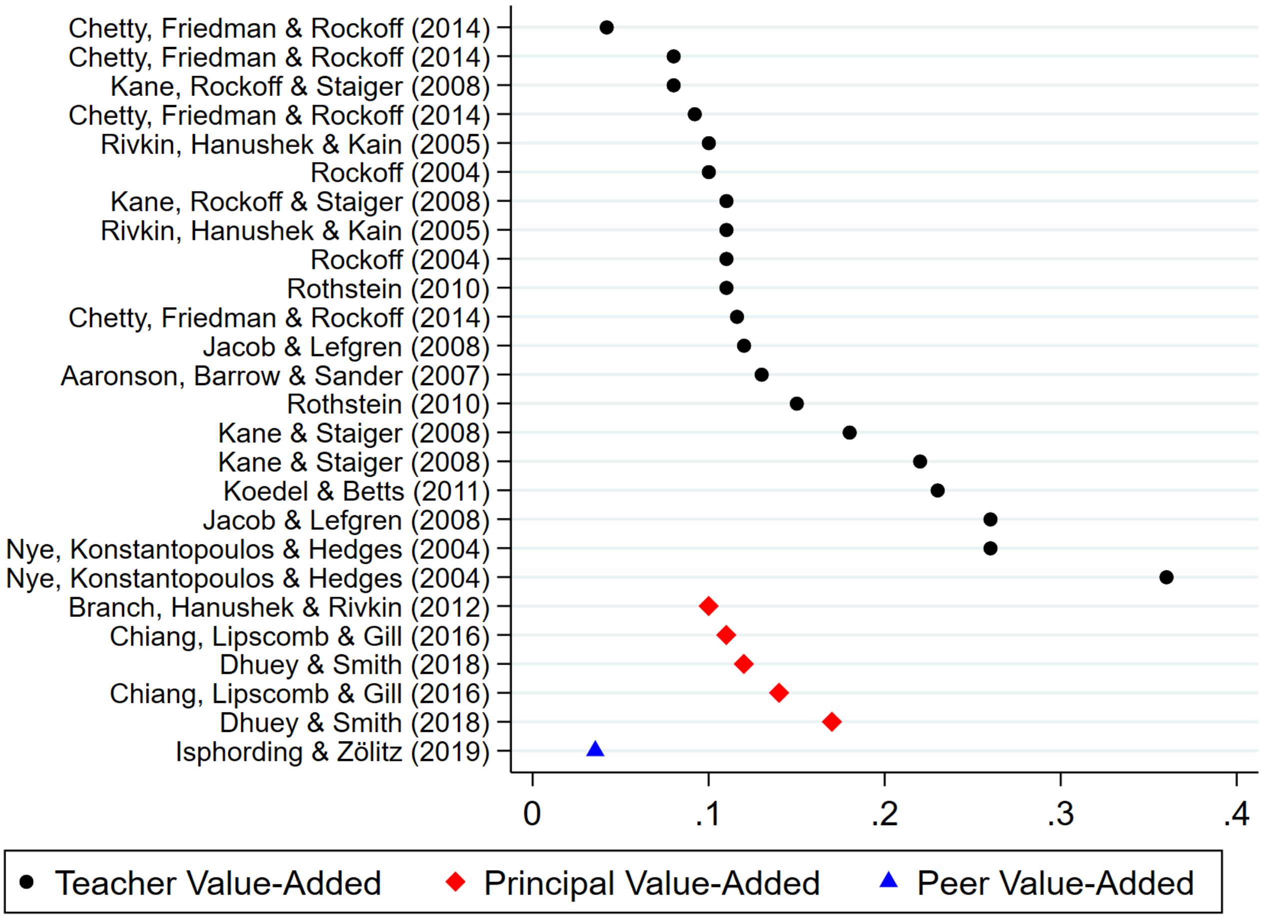
Note: Point estimates and confidence intervals shown are based on Table A3. The horizontal dashed line provides the full sample point estimate from our validation exercise for reference. The point estimates shown are based on one regression with a fully interacted model including instructor fixed effects. The *p*-value of low vs. medium teacher value-added = 0.005. The *p*-value of low vs. high teacher value-added = 0.0566.

Figure 8: Malleability of Peer Value-Added



Note: This bin scatter shows the relationship between standardized own PVA in second and third years and peer capital. Peer capital is the standardized average PVA of the peers an individual met in the first study year. Both peer value-added and peer capital are standardized to mean zero and standard deviation one. Figure 2 shows the distribution of peer capital at the student level. The point estimate and significance level are obtained from OLS regression. Peer value-added is constructed based on second- and third-year social interactions. The variable 'Peer Capital' is constructed based on a separate dataset of first-year student-peer pairs. One observation represents one student. N= 3,117. Significance levels indicated as * $p<0.1$, ** $p<0.05$, and *** $p<0.01$.

Figure 9: Comparison of Teacher, Principal, and Peer Value-Added



Note: This figure summarizes value-added estimates of teachers (black), principals (red) and peers (blue) drawn from recent published economic studies. Estimate sizes are reported in standard deviations of distributions of value-added on standardized performance measures.

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 Out-of-sample Prediction Sample

	(1) Peers' Peer Value-Added	(2) Peers' Peer Value-Added	(3) Peers' Peer Value-Added	(4) Peers' Peer Value-Added	(5) Peers' Peer Value-Added
Female	-0.000042				-0.000037
	(0.000)				(0.000)
GPA		-0.000034			-0.000027
		(0.000)			(0.000)
Dutch			0.000079		0.000046
			(0.000)		(0.000)
German				-0.000069	-0.000032
				(0.000)	(0.000)
Observations	312,240	312,240	312,240	312,240	312,240
R-squared	0.013	0.013	0.013	0.013	0.013

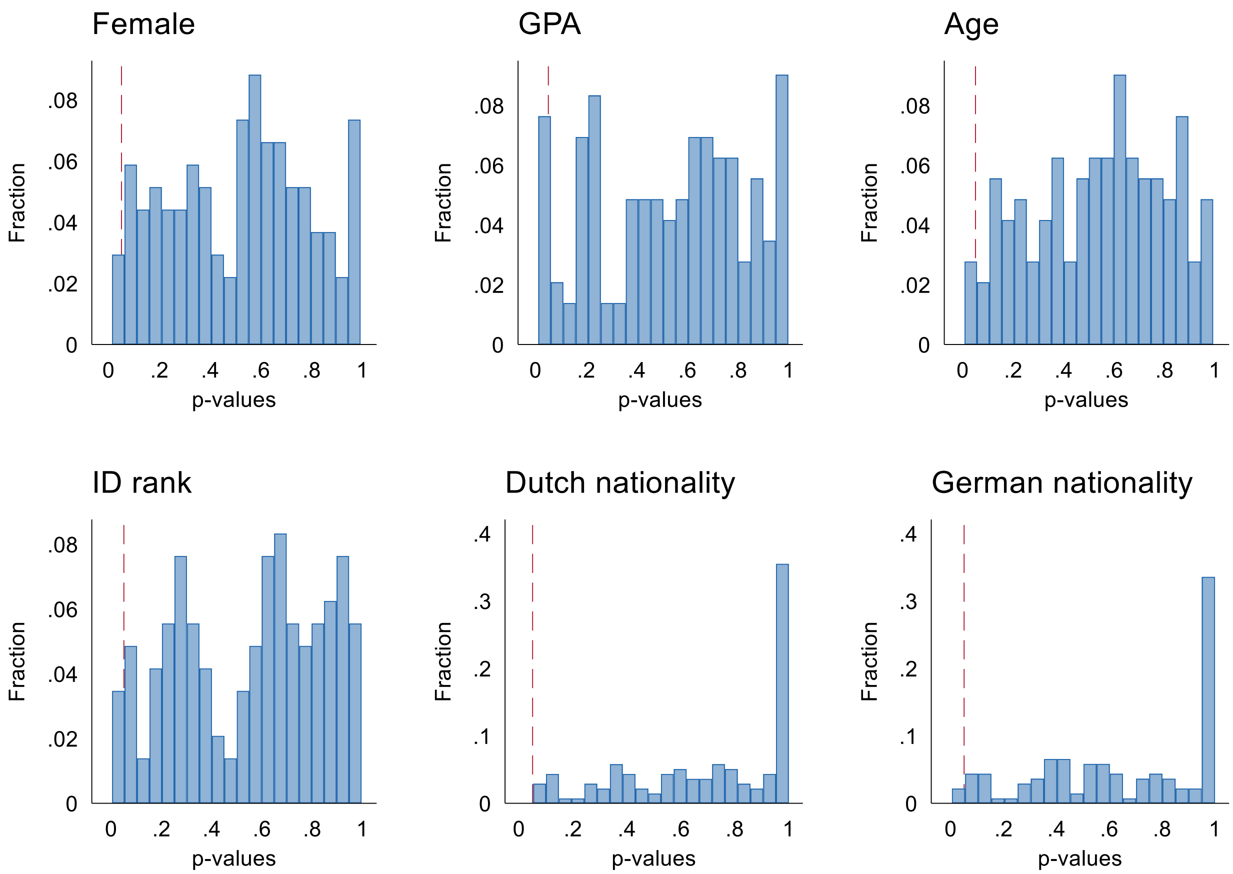
Note: This table reports results of regressions of PVA based on first-year social interactions on pre-determined student characteristics. One observation is one student-peer interaction. Robust standard errors using two-way clustering at the individual and section level are in parentheses. Significance levels indicated as * p<0.1, ** p<0.05, and *** p<0.01.

Table A3: Interaction between Peer Value-Added and Teacher Quality

	(1) Std. Course Grade	(2) Std. Course Grade	(3) Std. Course Grade	(4) Std. Course Grade
Peer Value-Added	0.0993***	0.0742**		
	(0.038)	(0.031)		
Peer Value-Added * Low Teacher Value-Added			0.1998***	0.2017***
			(0.057)	(0.058)
Peer Value-Added * Medium Teacher Value-Added			-0.0203	-0.0386
			(0.065)	(0.066)
Peer Value-Added * High Teacher Value-Added			0.0712	0.0220
			(0.071)	(0.065)
Observations	312,244	312,244	247,291	247,291
R-squared	0.450	0.459	0.450	0.458
Instructor fixed effects	No	Yes	No	Yes

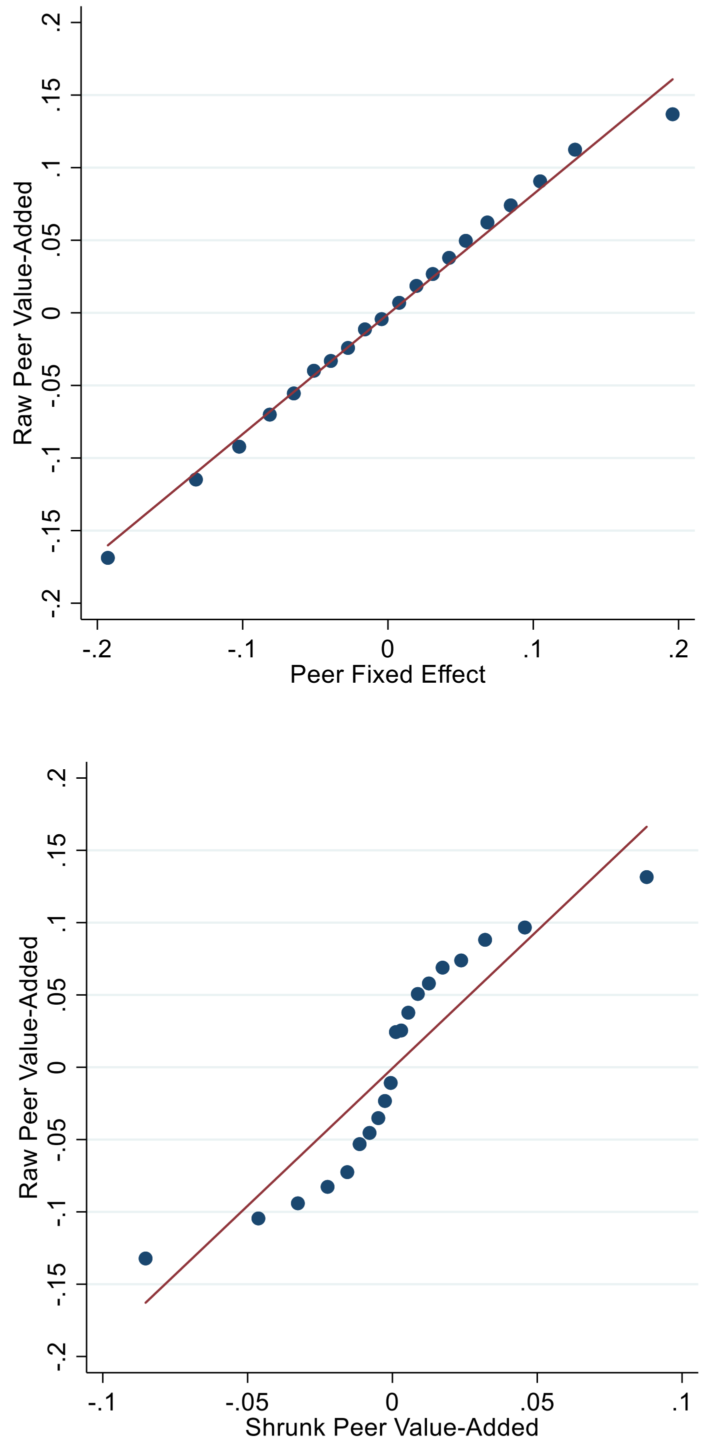
Note: This table reports results building upon the model we use for the validation exercise in Table 6. One observation per student-peer interaction. Figure 7 is based on estimation results shown in column (4). The number of observations in columns (3) and (4) is lower as we can only compute teacher value-added for instructors with an observable staff ID who teach multiple courses. When constructing teacher value-added, we exclude the course of interest. Columns (1) and (2) report robust standard errors using two-way clustering at the individual and section levels in parentheses. Columns (3) and (4) report robust standard errors using two-way clustering at the individual and instructor levels in parentheses. Significance levels indicated as * p<0.1, ** p<0.05, and *** p<0.01.

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



Note: These are histograms with 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 stratification of assignment based on nationality.

Figure A2: Correlation between Peer Value-Added and Peer Fixed Effect



Note: The bin scatter in the upper panel shows the relationship between raw PVA and peer fixed effect. The lower panel shows the relationship between shrunk and unshrunk PVA. Correlation coefficients are 0.9125 in the upper panel, 0.8138 in the lower panel, and significant at the 0.001 percent level.

Appendix A2: Robustness Check Regarding Indirect Reflection Problem

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 individual student enters the estimation *either* as peer or student, but never in both roles simultaneously.

Computing PVA in this way provides results comparable to the main specification. PVA remains predictive in our out-of-sample validation exercise among newly assigned students. The coefficient we obtain when we estimate the model comparable to Table 6, column (1) is 0.11 and significant at the 5 percent level. We are therefore not concerned about simultaneity affecting our PVA estimates.