Teacher Peers at School: How Do Colleagues Affect Value-Added and Student Assignments?

Hedvig Horváth*

THIS VERSION: October 31, 2016

Please click here for the most recent version.

Abstract

I estimate the effect of peers, defined as teachers at the same school and grade level, on the own effectiveness of teachers, measured by value-added scores. Traditional estimates using leave-out means imply significant and large positive spillovers among teachers. In this paper, I exploit a more compelling research design following Mas and Moretti (2009), which approximates the following thought experiment in a regression framework: A low value-added teacher, Teacher B is randomly replaced by a new, high value-added teacher, Teacher C, at a particular school-grade level. How does the value-added of incumbent Teacher A, who worked at the same school-grade level the year before, change in response? Among North Carolina elementary school teachers, I find that in response to the replacement of a teacher by a 1 standard deviation (SD) better teacher, incumbent peers' value-added increases by about 0.12 SDs, which is about one half of traditional estimates. Moreover, I uncover large heterogeneities across schools, which shed light on how much of these estimates are due to pure peer effects among teachers versus student sorting. Using the classification of schools developed in Horváth (2015), I find that the results are driven by schools that sort students to teachers, while estimates are low and insignificant in random assignment schools. Looking at changes in observed student characteristics in the incumbent teachers' classrooms reveals that assignments in random schools do not change when a colleague is replaced, just as expected. Therefore, the insignificantly small estimates in these schools provide clean estimates of peer effects. In contrast, average prior achievement in incumbent teachers' classrooms significantly decreases in sorting schools. In these schools, therefore, student sorting may explain the spillovers among teachers.

* University College London, Institute of Education,h.horvath@ucl.ac.uk. I would like to thank David Card, Pedro Carneiro, Bruce Fuller, Zoltán Hermann, Patrick Kline, Attila Lindner, Jesse Rothstein, Uta Schönberg, David Silver, George Stoye for their invaluable advice. Participants at the Applied Lunch and IOE-QSS seminars at UCL, CERS-HAS Summer Workshop and CReaM's Topics in Labor Economics Workshop provided excellent comments. All remaining errors are solely mine.

1 Introduction

Imagine that Teacher A teaches at fourth grade in a school for at least two consecutive years, while her colleague, also teaching fourth graders at the same school, gets replaced by a 1 standard deviation (SD) better teacher from one year to the next. How does the achievement of Teacher A's students change in response to the replacement of colleagues? This thought experiment mimics the set-up of Mas and Moretti (2009) for supermarket cashiers. Provided that the replacement occurs randomly and on average similar students are assigned to Teacher A irrespective of who her colleagues are, we can interpret the answer to the previous question as the peer effect among school teachers.

There are several reasons to think that (positive) spillovers are not negligible among teachers. Teachers facing similar students may share successful teaching practices (*peer learning*), an effective peer may incentivize his/her colleagues to exert more effort (*social pressure*), or students of different teachers at the same school-grade level may share knowledge (*student interaction*). However, the potential endogeneity of the replacement of colleagues casts doubts on this interpretation, and a nonzero Mas and Moretti type estimate may give rise to alternative explanations as well. First, *unobserved shocks* may coincide with the replacement of teachers (e.g. change in the student body, introduction of a new teaching technology, etc.), and second, the quality of students assigned to the incumbent teachers may change in response to the replacement (*sorting effect*).

In this paper, I approximate the above thought experiment in a regression framework. I adopt the approach of Mas and Moretti (2009) and estimate the effect of peer replacement on the value-added score of incumbents among elementary school teachers in North Carolina. Next, I take advantage of the classification of schools developed in Horváth (2015), to disentangle this effect into pure peer effects among teachers and student-to-teacher sorting effects.

This paper relates to two important branches of previous research: to the (i) methodologically oriented teacher value-added estimation literature, but more importantly, to the

(ii) research on peer effects, in particular, among teachers. Aaronson et al. (2007) find large effects of the English teacher on math scores for Chicago Public School 9th graders, however, they leave it for further research to decide whether it is due to sorting or teacher peer effects. Two papers pursue that. A field experiment conducted by Papay et al. (2016) suggests that peer learning works among teachers: the authors find a persistent, 0.12 SD increase in the value-added of previously low-performing teachers after being paired up for consultation with a high-performing peer for a school year. Jackson and Bruegmann (2009) also argue for a peer learning mechanism explaining the spillovers they find for North Carolina teachers. They estimate teacher value-added as fixed effects from 1996-2000 North Carolina data and then form peer value-added as the average of value-added scores of teachers teaching at the same school-grade level but not teaching a given teachers students by construction. Afterwards, they regress student achievement on both own- and peer teacher value-added scores on 2001-2006 data. They find that 1 SD better peers increase own students' scores by 0.03-0.04 SDs.

In this paper, I exploit similar identifying variation as Jackson and Bruegmann (2009) do: teacher switches across grade-levels and schools in North Carolina. By adopting the approach of Mas and Moretti (2009), I build on their appoach but arrive at different conclusions. First, by looking at the effects of peers on incumbent teachers only, in light of Angrist (2014), I strictly separate research subjects (incumbent teachers in a given year) from treatment peers (teachers who enter and exit from the given school-grade cell). Second, instead of splitting the time period into two spells, I construct the measure for peer quality as leave-two-years-out shrunken value-added (Kane and Staiger (2008), Jacob et al. (2010), Chetty et al. (2014a)) to minimize measurement error and the need for imputation. Similarly to Jackson and Bruegmann (2009), I also compute teacher value-added in a first stage, which assumes only own-teacher affect student achievement, and create peer value-added using these estimates in a second stage.[1] However and more importantly, I specify the second-stage

[1]Work in progress improves on this and adopts a specification that allows for teacher peer effects to exist even in this first stage.

econometric model in changes instead of levels, which mitigates the reflection problem arising from ignoring potential peer effects in the first stage (Manski (1993)). Augmenting the model in changes with school-year school grade effects controls for additional trend heterogeneities across incumbent teachers. Finally and most importantly, I compare the estimates in schools where students are randomly assigned to teachers (random assignment schools) and in other schools, where students are tracked into different classrooms based on prior achievement, and certain teachers, year after year, teach the high-achieving classrooms, while other teachers always teach the low-achieving classrooms (tracking & matching schools).[2] As sorting effects are most likely zero in random assignment schools, Mas and Moretti type estimates in such schools can be interpreted as pure spillover effects. In tracking & matching schools, however, sorting effects may dominate and peer effects may be small. First, in such schools there may be less room for peer learning or social pressure since teachers face very different students. Second, the arrival of a new teacher may change which students are assigned to which teachers, and these assignments may be reflected in teacher value-added scores.

I find that a 1 SD increase in peer value-added results in about a 0.12 SD increase in own value-added, highly statistically significant. The magnitude is about one half of traditional, leave-out mean estimates, but more than twice as large as those found by Jackson and Bruegmann (2009). However, these results are driven by tracking & matching schools, while estimates in random assignment schools are much more moderate (0.04 SDs in response to a 1 SD increase in peer value-added). These patterns suggest that the Mas and Moretti type estimates are dominated by sorting effects occurring in tracking & matching schools, where the entry of a new teacher changes which students are assigned to which teachers. In contrast, cleaner estimates of peer effects, identified as Mas and Moretti type estimates in random assignment schools, are small. This is also supported by the finding that average prior achievement in incumbent teachers' classrooms decrease, but only in tracking & matching schools.

[2]This classification of schools is described in more details in Horváth (2015).

This last point relates to the other strand of literature – about teacher value-added estimation – that this paper speaks to. Kane and Staiger (2008) exploited a small-scale teacher switching experiment in Los Angeles to confirm that observational estimates of teacher value-added are not biased by nonrandom sorting of students to teachers, however, their results are noisy. Kane et al. (2013), Chetty et al. (2014a) and more recently Bacher-Hicks et al. (2014) replicated the underlying thought experiment in different school districts using teacher switches between schools and grade levels as quasi-experiments and arrived at the same conclusion. However, Rothstein (2015) pointed out that the quasi-experimental nature of these teacher switches may be undermined, because in North Carolina he finds evidence for changes in predicted teacher value-added at the school-grade level positively correlating with changes in average prior achievement at the school-grade level. The results I find in tracking & matching schools cast doubt on a different underlying assumption of these quasi-experiments: When a new teacher enters, the assignment of the incumbent teachers also changes in sorting schools.

However, the exact mechanism behind this last result is subject to further research, it may have important policy implications. If the arrival of a new teacher affects the assignment and the value-added of her incumbent peers, school districts should be cautious using such value-added measures as a basis for individual teacher evaluation such as firing and tenure decisions or financial remuneration. With such a policy, a bad teacher may punish other teachers as well, and thus teachers may also cluster across schools by quality, affecting educational inequalities.

The rest of the paper is organized as follows. Section 2 briefly presents a test score production function framework that distinguishes between pure peer effects among teachers and student sorting effects. In Section 3, I describe the empirical methods, namely, the estimation procedure for teacher value-added and the adoption of framework of Mas and Moretti (2009) for teachers. Section 4 presents the results and discusses the implications, while Section 5 concludes and outlines directions for further research.

2 Framework: Peer- vs. Sorting Effects in the Test Score Production

Assume that student i's end-of-grade test score, y_{it}, is determined by the following data generating process:

$$y_{it} = X_{it}^{'}\beta + Z_{it}^{'}\rho + \mu_j + \mu_{-jt} + \varepsilon_{it}, \tag{1}$$

where X_{it} is a vector of the observed characteristics of student i at school-grade-year cell sgt, while Z_{it} is the vector of characteristics unobserved to the researcher but potentially observed by the principal. μ_j is teacher j's value-added, while $-jt$ denotes her year t coworkers' value-added.[3,4,5] ε_{it} is an individual shock with $\varepsilon_{it} \sim iid\, N\left(0, \sigma_{\varepsilon}^2\right)$.

Since Z is unobserved to the researcher, in practice we estimate

$$y_{it} = X_{it}^{'}\tilde{\beta} + \mu_j + \mu_{-jt} + \tilde{\varepsilon}_{it}, \tag{2}$$

where

$$\tilde{\varepsilon}_{it} = \tilde{Z}_{it}^{'}\rho + \varepsilon_{it},$$

and $\tilde{Z}$ comes from the projection $Z = X\kappa + \tilde{Z}$, and so $\tilde{\beta} = \beta + \kappa\rho$.

Let

$$\nu_{it} = \tilde{Z}_{it}^{'}\rho + \mu_j + \mu_{-jt} + \varepsilon_{it}.$$

Then averaging over all of j's students in year t yields

$$\bar{\nu}_{jt} = \overline{\tilde{Z}}_{jt}^{'}\rho + \mu_j + \mu_{-jt} + \bar{\varepsilon}_{jt}.$$

[3] j denotes teacher assignment $j = j(i,t)$ if student i in year t but notation is reduced for easier reading.

[4] The definition of coworker in the context of teachers will be specified later in Section XXX.

[5] Instead of the additively separable specification in own- and peer teacher effects, one could allow the test score to depend on a flexible function of own-teacher and coworker-teacher effects, $\mathrm{m}\left(\mu_j, \mu_{-jt}\right)$ in (1). In that case, however, obtaining an estimate for individual teacher effects is more challanging. See Silver (2016) for the details of this approach in the physician context.

After first differencing, we get

$$\triangle\bar{\nu}_{jt} = \triangle\overline{\tilde{Z}}'_{jt}\rho + \triangle\mu_{-jt} + \triangle\bar{\varepsilon}_{jt}.$$

When estimating the effect of peers on a teacher's value-added, ideally, we would regress $\triangle\bar{\nu}_{jt}$ on $\triangle\overline{\tilde{Z}}'_{jt}$ and $\triangle\mu_{-jt}$, and the parameter of interest would be the the coefficient sitting on $\triangle\mu_{-jt}$. However, since $\tilde{Z}$ is unobserved and we only have an estimate of $\triangle\mu_{-jt}$, what is feasible is to

$$\text{regress } \triangle\bar{\nu}_{jt} \text{ on } \triangle\hat{\mu}_{-jt}.$$

Then the coefficient sitting on $\triangle\hat{\mu}_{-jt}$ will contain omitted variable bias,

$$\frac{\rho\mathrm{Cov}\left(\triangle\overline{\tilde{Z}}'_{jt}, \triangle\hat{\mu}_{-jt}\right)}{\mathrm{Var}\left(\triangle\hat{\mu}_{-jt}\right)}. \tag{3}$$

I assume that the pool of students in a school-grade cell is exogenously given, and the school principal is in full charge of assigning students to teachers. In some schools, this assignment process is random. In other, so called "tracking & matching" schools, students of similar prior achievement are tracked and certain teachers, year after year, get the high-achieving classroom, while other teachers always get the low-achieving classrooms. Horváth (2015) provides a statistical procedure to classify schools into these two categories. $\overline{\tilde{Z}}'_{jt}$ is assumed to be 0 in random assignment schools but not in tracking & matching schools. Therefore, (3) omitted variable bias is 0 in random assignment schools but potentially non-zero in tracking & matching schools. The sign depends on how the assignment of incumbents change when a better teacher replaces an average one: if high (low) $\tilde{Z}$, that is, high (low) potential gain students get assigned to the new peer, the bias will be positive (negative), as the incumbent teacher will receive lower (higher) potential gain students. Hence, two testable implications follow:

1. In random schools, neither assignments nor own value-added of incumbent teachers are

affected by changes in peer value-added.

2. In tracking & matching schools, if a teacher is replaced by a better teacher, the value-added and the student assignments of the incumbent teachers may change. Hence, in these schools, peer effect estimates may be contaminated by a sorting bias, (3). The bias is positive (negative) if high (low) potential gain students get assigned to the new peer and the incumbent teacher receives lower (higher) potential gain students.

3 Empirical Methods

3.1 Estimating Teacher Value-Added

I estimate the following student-level value-added model (VAM):

$$y_{it} = X_{it}'\beta + \nu_{it}, \tag{4}$$

where in all specifications the dependent variable is the standardized end-of-grade math test scores of student i, taught by teacher $j = J(i,t)$ at school $s = S(i,t)$ in grade $g = G(i,t)$ at year t, y_{it}, and controls, X_{it} include

- grade-specific, 3rd order polynomial of lagged scores, both math and reading;
- student's gender, ethnicity, exceptional status, FRL status and parental education;
- year$\times$grade dummies.

To get the teacher effect estimates, $VA_{jt}^{-\{t,t-1\}}$, after estimating the student-level model, I apply the method described in Chetty et al. (2014a). This improves on the approach used by Kane and Staiger (2008) and Jacob et al. (2010) by allowing, more flexibly, for a drift in true value-added. Hence, we get teacher value-added scores:[6]

[6]The estimation is implemented using the "vam" STATA code written for Chetty et al. (2014a).

$$\widehat{VA}_{jt}^{-\{t-1,t\}} = \sum_{r \neq 0,1} \psi_{|r|} \bar{\nu}_{j,t-r}, \tag{5}$$

$$\widehat{VA}_{jt}^{-\{t,t+1\}} = \sum_{r \neq 0,-1} \psi_{|r|} \bar{\nu}_{j,t-r}, \tag{6}$$

where $\bar{\nu}_{j,t-r}$ is the average of regression residuals in (4) for teacher j's students in year $t-r$. $\psi_{|r|}$'s are OLS coefficients from regressing $\bar{\nu}_{jt}$ on $\{\bar{\nu}_{j,t-r}\}_{r \neq 0,\pm 1}$. These only depend on how far apart average residuals, $\bar{\nu}_{j,t-r}$, is from the year for which we want to estimate value-added (t). By construction, these value-added estimates are best linear predictors for $\bar{\nu}_{jt}$, that is, in the OLS of

$$\bar{\nu}_{jt} = \lambda_0 + \lambda_1 \widehat{VA}_{jt}^{-\{t,t\pm 1\}} + \xi_{jt}, \tag{7}$$

we expect λ_1 to be statistically insignificantly different from 1, (and λ_0 to be insignificantly different from 0). The results are displayed in column (1) of Table 1 and show that the estimates slightly overshrink the series of average residuals, as λ_1 is significantly larger than 1, even if the point estimate is only slightly higher.[7,8]

3.2 Adopting the Framework of Mas and Moretti (2009) for Teachers

After computing leave-2-out value added measures for each teacher, I define the change in the peer value added, $\triangle VA_{-jgst}$, as

$$\triangle\overline{VA}_{-jgst} = \frac{\sum_{k \text{ new}} \widehat{VA}_{kgst}^{-\{t-1,t\}}}{\text{number new}} - \frac{\sum_{k \text{ leaving}} \widehat{VA}_{kgs,t-1}^{-\{t-1,t\}}}{\text{number leaving}}. \tag{8}$$

[7]The point estimates for the regression with the leave-1-year-out value-added on the right handside has the same coefficients as in Rothstein (2015).

[8]In robustness checks (not reported), all results in the paper are replicated using teacher value-added scores estimated in the Kane and Staiger (2008) fashion. These value-added scores undershrink the series of average residuals (forecast coefficient 0.972 with s.e. 0.004). Using the alternative shrinkage, results in the paper would be qualitatively the same but somewhat smaller due to the slight overshrinkage here and the undershrinkage with the alternative method.

A couple of things deserve to be noted. This measure is nonzero only if a new teacher comes in or a teacher leaves. The better the new teacher is, the higher $\triangle\overline{VA}_{-jgst}$ is, or the better the leaver is, the lower $\triangle\overline{VA}_{-jgst}$ is. As Angrist (2014) points out, when estimating peer effects, one should separate research subjects from treatment peers, therefore, I do not include incumbent peers' value-added in (8). I do not weight by class sizes, as those may be endogenous.

I run the following regression:

$$\triangle\bar{\nu}_{jgst} = \varrho_0 + \varrho_1\triangle\overline{VA}_{-jgst} + Z^{'}_{gst}\psi + \varepsilon_{jgst}, \tag{9}$$

where Z_{gst} is the change in the number of peers, change in the pool of students at the school-grade-year level (change in the relative class size, fraction of nonwhite students, and average lagged test scores). Because change in peers' value-added appears on the right handside, I use the shrunken value-added scores in (8) to provide a value-added score with the least measurement error.[9] I estimate eq. (9) for incumbent teachers who will stay for at least another year. This latter is important to make sure that no teacher-year observation that appears as a treatment peer shows up as a research subject (Angrist (2014)). In some specifications, I also add principal, school-grade or school-year effects to control for unobserved effects related to the person in charge of teacher assignments, or to capture unobserved shocks.[10]

The advantage of the change specification, as opposed to levels, is that it mitigates the reflection problem that arises due to the fact that future years are also used to compute peer value-added scores, and as such, these may be affected by teacher j through peer effects. If teacher j influences her future colleagues to a similar extent than her past colleagues, this influence is cancelled out in the change specification, even if potential peer effects were not

[9] As outlined in Section 3.1 and similarly to Chetty et al. (2014a), I use a leave-2-years-out shrinkage estimate for value-added on the right handside, so that whatever is used to generate this estimate does not contaminate mechanically the value-added on the left handside.

[10] Note that because the peer treatment is the same for everyone in the estimation sample at the same school-grade-year cell, it is not possible to control for school-grade-year effects. Thus, in principle, it is plausible (though unlikely) that the coefficients I find reflect unobserved school-grade level shocks exactly in years when a better teacher enters.

taken into account in the first stage, teacher value-added score estimation (in eq. (4)).

Our parameter of interest is ϱ_1, which tells us that if a 1 standard deviation (SD) better teacher arrives at the school-grade level, incumbents' value-added changes by ϱ_1 SDs on average.

To see if the observed characteristics of students assigned to the incumbent teachers change, I replace the change the left handside variable in eq. (9) with changes in teacher j's students' characteristics (such as relative class size, relative fraction of nonwhite students in the classroom[11] and average lagged achievement in the classroom).

Finally, to see if the effect of peers is different in tracking & matching schools than in random schools, I add the main effect and the interaction term for the assignment policy variable (a categorical variable for tracking only and tracking & matching) identified in Horváth (2015):

$$\begin{aligned} \triangle\bar{\nu}_{jgst} &= \varrho_0^0 + \varrho_0^1 T_s + \varrho_0^1 TM_s + \\ &+ \varrho_1^0 \triangle\overline{VA}_{-jgst} + \varrho_1^1 \triangle\overline{VA}_{-jgst} T_s + \varrho_1^2 \triangle\overline{VA}_{-jgst} TM_s + \\ &+ Z_{gst}^{'}\psi + \varepsilon_{jgst}, \end{aligned} \tag{10}$$

3.3 Data and Summary Statistics

I use the administrative dataset linking 3rd-5th grade students to their teachers between 1995 and 2013, provided by the North Carolina Education Research Data Center. At these grade levels, classrooms are mostly taught by one single teacher and every year, students get assigned to new teacher (and classmates). Teachers can be followed from one school to another as long as they stay within the public K-12 school system of North Carolina, and their value-added scores can be computed using their students' end-of-grade math and reading scores. The results in this paper are presented for math value-added scores only,

[11] Relative class size is defined as $\frac{\text{class size}}{\text{number of students at school-grade-year cell}}$, while the relative fraction of nonwhite students as $\frac{\text{nonwhite students in class}}{\text{nonwhite students at school-grade-year cell}}$.

which are standardized to have mean 0, sd 1 in each year-grade. In the student-level VAM, I use the following controls: race, gender, homework, TV in all years; parental education, FRL and exceptionality status, days absent in selected years. Peers ("colleagues") are defined as teachers teaching a different classroom at the same school, grade level in the same year.

Summary statistics of the main variables that I use in this paper are displayed in Table 3 and histograms in Figure 1. These are the following: average classroom residuals, shrunken value-added computed as described in Section 3.1 above, change in average classroom residuals and in peer value-added for stayers.[12]

Tables 2 shows that the number of teacher at a school-grade-year cell is highly right skewed; on average there is one teacher leaving and one arriving at the school-grade level, with 2 additional teachers staying. Turnover is slightly higher in tracking & matching schools. Table 3 shows generally that value-added of leavers are the lowest, while that of the stayers is the highest, across all school types. However, average classroom residuals are the lowest for newcomers, that may reflect a temporary period of adaption to the new environment, or an initial unfavorable assignment. The histograms reveal that while the distribution of change in average classroom residuals is almost identical across school random and tracking & matching schools,[13] in random schools there changes in peer value-added around zero are much more common, only due to the fact that teacher switches occur more frequently in tracking & matching schools (and this is when the change in peer value-added is nonzero).

4 Results and Discussion

As a benchmark for traditional peer effect regressions, I estimated a regression similar to eq. (7), but with the leave-out mean of peers' value-added included:

$$\bar{\nu}_{jt} = \lambda_0 + \lambda_1 \widehat{VA}_{jt} + \lambda_2 \overline{\widehat{VA}}_{-jt} + \xi_{jt}.$$

[12] Again, peer value-added is defined as the difference of value-added between newcomers and teachers who left since last year from the school-grade level.

[13] These school types were identified in Horváth (2015).

Although Angrist (2014) talks about the flaws of interpretation of the estimate for λ_2 in such a model, it may be a useful comparison to main results with a better design. Results are shown in columns (2) of (1). If I did not have the better design, then these estimates would suggest much larger peer effects (0.26 SDs) than what my preferred estimates will show later.

Main results from the Mas and Moretti design are shown in Tables 4, 5 and in Figure 2. Columns (1) through (3) show estimates for all schools, while columns (4) through (6) apply for school-grade cells that are not expanding or contracting, that is, the number of teachers does not change.

Table 4 shows the baseline results and Panel A of Figure 2 graphically illustrate the preferred specification (with school-grade and school-year effects). There is a strongly positive association between the change in non-incumbent peer value-added and own value-added. A 1 standard deviation (SD) increase in peer value-added results in a 0.10-0.12 SDs increase in own value-added. Controls for observed average characteristics of the student body at the school-grade-year level (change in number of students, fraction of nonwhite students, average prior achievement) essentially do not change the results. This means that changes in teacher quality is uncorrelated with other changes in school-grade-year level observables. However, when adding principal-, school-year or school-grade effects, the coefficients decrease, meaning that higher value-added teachers enter in schools where incumbent value-added is also increasing. Nevertheless, even after controlling for these unobserved heterogeneities and restricting the sample to school-grade cells with a constant number of teachers, the positive association between changes in peer- and own value-added remains.

Beyond the overall results, looking at heterogeneities by the sorting type of schools sheds light on the mechanism behind these positive effects. Therefore, I next look at differences in the peer effect between random and tracking & matching schools, classified in Horváth (2015). Recall that in tracking & matching schools, students in the same school-grade-year cohort are grouped into homogenous classrooms based on their prior test scores, and some teachers year after year get assigned to the high-achieving classrooms, while other teachers to the low-achieving classrooms. In random schools, neither of these two occurs.

Regression results showing the difference between the two types of schools are displayed in Table 5, and graphically in Panel B of Figure 2. A striking pattern emerges from Table 5: The magnitude of changes in peer value-added on own value-added is similar as in the overall specifications, however, for all the schools, the results are driven by random schools, while when we restrict the sample to non-expanding and non-contracting schools (Column 4), they are driven by tracking & matching schools. In random schools, assignments should not (and as we will see in the next tables, do not) depend on peer value-added, so the coefficients for random schools are estimates for a combination of peer effects and effects of unobserved shocks. The fact that the coefficient decreases to 0 when we restrict the sample to non-expanding/non-contracting school-grade-year cells and include school-grade and school-year effects to control for unobserved heterogeneities, suggests that the positive estimates mostly reflected other specificities at the school-grade cell at the same time when the high value-added peers entered.

On the other hand, the additional effect in tracking & matching schools estimate the net effect of peer effects and sorting effects. If anything, peer effects are expected to be weaker in tracking & matching schools, since teachers teach different types of kids, and so there is less room for peer learning. The fact that in non-expanding/non-contracting schools, the coefficient on peer value-added is only significantly positive at tracking & matching schools and much smaller, insignificant positive at random schools suggests that sorting is more likely to be the underlying mechanism. This is illustrated graphically in Panel B of 2: peer value-added has a visibly larger effect on own student achievement in sorting schools.

The above results suggest that when high value-added peers move in, sorting of students to incumbent teachers may significantly change, at least in tracking & matching schools. Therefore, Tables 6-8 explore how differently the student assignments of incumbent teachers change in the two school types. Only results for non-expanding/non-contracting school-grade level cells are shown. These are the preferred specifications, as in schools where the number of teaching staff increases/decreases in year, there are likely to be other changes as well, which might invalidate the design. Effects on log own class size are either precise zeros, or economically tiny negatives in all specifications. That is, principals do not adjust class sizes when the quality distribution of teachers shift. Fraction of nonwhite students in own classroom of incumbent teachers does not change either.

However, average prior achievement in incumbent teachers' classrooms decreases in response to a high value-added teacher entering the same school-grade level. If a teacher is replaced by a 1 SD better colleague, average lagged math scores in an incumbent teacher's classroom decrease by 0.04-0.06 SDs (0.01-0.06 for reading).[14] Moreover, this negative effect can fully be attributed to tracking & matching schools; incumbent teacher assignments are not affected in random schools. Interestingly, in the regression bunching together all school types, better peers are associated with even lower prior achievement in a teacher's own classroom (not reported) when controls for the student composition at the school-grade level cells are added. This is in line Rothstein (2015)'s point that better teachers enter in school-grade level cells where prior achievement increases anyways.

Overall, these results suggest that the strong positive association between changes in peer- and own value-added comes from expanding random schools (0.10-0.12 SDs; most likely due to other coinciding unobserved shocks at the school-grade level) and tracking & matching schools where the size of the teaching staff is constant (0.16 SDs). This latter reflects changes in the student assignments of incumbent teachers – a decline in the average prior achievement in their classrooms.

[14]Evidence for prior reading achievement (not reported) in incumbent teachers' classrooms is weaker, which may be due to the fact that all value-added measures are computed from math test scores.

5 Concluding Remarks and Directions for Further Research

When a teacher is replaced at a school-grade level by a better teacher, the value-added of her incumbent coworkers rises. In this paper, I argued that this rise is mainly due to changes in the student assignment at the grade-level or other unobserved shocks, and less so to peer effects among teachers. When a high value-added teacher enters, class size of the incumbents does not change significantly. However, in tracking & matching schools, incumbents tend to be assigned to lower prior achievement kids. In random schools, peer value-added still positively affects own value-added, although this is driven by expanding schools, and assignments do not change.

A couple of issues and related questions remain for further research. First, why do incumbent teachers get assigned to lower prior achievement kids when a high value-added colleague enters? A model where the all students have to achieve at grade level would suggest the opposite, while common sense would suggest that if a high value-added teacher shows up, she should get the favorable (high prior achievement) classroom, while incumbents get observationally worse students. If so, however, why does the value-added of incumbents increase at the same time, exactly at schools where the assignments change? Results in Horváth (2015) suggests that lower value-added teachers get lower prior achievement students, however, these are in fact the high-gain students, and as a consequence, value-added of these teachers will be upward biased. This story is consistent with the findings here, however, it remains a puzzle why low prior achievement students would be the high gain ones. This needs further exploration.

A couple of related questions remain. The first is to look at if the changes in own value-added in response to a change in peer value-added are permanent. Second, are these effects symmetric? The arrival of a bad teacher is unlikely to destroy her incumbent colleagues teaching effectiveness, so if results hold for the replacement of a good teacher with a bad one just firmly as for the opposite, that is additional evidence for the dominance of sorting effects.

References

Aaronson, D., L. Barrow, and W. Sander, "Teachers and Student Achievement in the Chicago Public High Schools," *Journal of Labor Economics*, 2007.

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

Bacher-Hicks, A., T. J. Kane, and D. O. Staiger, "Validating Teacher Effect Estimates Using Changes in Teacher Assignments in Los Angeles," 2014. NBER Working Paper No. 20657.

Chetty, R., J. N. Friedman, and J. E. Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 2014a.

Horváth, H., "Classroom Assignment Policies and Implications for Teacher Value-Added Estimation," January 2015.

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

Jacob, B. A., L. Lefgren, and D. Sims, "The Persistence of Teacher-Induced Learning Gains," *Journal of Human Resources*, 2010, pp. 915–943.

Kane, T. J. and D. O. Staiger, "Estimating Teacher Impacts On Student Achievement: An Experimental Evaluation," 2008. NBER Working Paper No. 14607.

_ , D. F. McCaffrey, T. Miller, and D. O. Staiger, "Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment," Technical Report, Bill & Melinda Gates Foundation 2013.

Manski, C. F., "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 1993, *60* (3), 531–542.

Mas, A. and E. Moretti, "Peers at Work," *American Economic Review*, March 2009, *99* (1), 112–145.

Papay, J. P., E. S. Taylor, J. H. Tyler, and M. Laski, "Learning Job Skills from Colleagues at Work: Evidence from a Field Experiment Using Teacher Performance Data," 2016. NBER Working Paper No. 21986.

Rothstein, J., "Revisiting the Impacts of Teachers," 2015. Working Paper.

Silver, D., "Haste or Waste? Peer Pressure and the Distribution of Marginal Returns to Health Care," 2016. Job Market Paper.

Tables

Table 1: Value-Added Scores as Best Predictors, Chetty et al. (2014a) method

	(1)	(2)
VARIABLES	avg. resids	avg. resids
leave-1-out VA	1.018	0.941
	(0.004)	(0.004)
peer VA		0.260
		(0.007)
Constant	-0.011	-0.012
	(0.001)	(0.001)
Observations	167,249	159,009
R-squared	0.316	0.328
sample	all	all
controls	no	no
FEs	no	no

Note: Standard errors in parentheses are clustered at the school level. Regressions are weighted by class size. Dependent variable in each column is average classroom residuals from the given student-level VAM specification, while the independent variables are precision-weighted average classroom residuals for all but the current (leave-1-out), all but the current and previous (leave-2-out lag) or all but the current and next (leave-2-out fw) years available for the teacher in the given school and grade level, shrunken to adjust for measurement error. See more details on the shrinkage in Section 3.1. If value-added in other years are best linear unbiased predictors (BLUPs) for this year's value-added, we should see coefficients of 1 (and intercepts of 0). Here, coefficients are somewhat below 1, indicating undershrinkage, which may come from the fact that this shrinkage method assumes that value-added from any year are equally predictive of current value-added. In columns (3) and (4) the leave-out average value-added of teachers (not weighted by class size) who teach at the same school-grade in the given year are added as a benchmark of traditional peer effect regressions.

Table 2: Summary Statistics - School-Grade-Year Cells

		mean	sd	p10	p50	p90	max	N
number of all teachers	all schools	3.119	1.483	1	3	5	16	48,979
	R	2.956	1.388	1	3	5	14	21,392
	TM	3.381	1.557	2	3	5	16	22,885
number of new teachers	all schools	0.957	1.118	0	1	2	13	48,979
	R	0.863	1.050	0	1	2	11	21,392
	TM	1.054	1.780	0	1	3	13	22,885
number of stayers from last year	all schools	1.771	1.432	0	2	4	11	48,979
	R	1.711	1.376	0	2	4	10	21,392
	TM	1.917	1.510	0	2	4	10	22,885
number of teachers leaving next year	all schools	0.918	1.049	0	1	2	12	48,979
	R	0.8354	0.9915	0	1	2	12	21,392
	TM	1.003	1.104	0	1	2	10	22,885

Note: Chetty et al. (2014a) shrinkage applied.

Table 3: Summary Statistics - Teachers

		mean	sd	VA mean	VA sd	N
new teachers	all schools	1.005	1.146	0.002	0.113	52,375
	R	0.861	1.039	0.006	0.114	14,888
	TM	1.100	1.203	0.001	0.112	33,190
teachers leaving next year	all schools	1.000	1.119	-0.000	0.113	52,143
	R	0.879	1.041	0.004	0.116	15,210
	TM	1.082	1.166	-0.002	0.112	32,645
stayers from last year	all schools	1.781	1.462	0.015	0.142	92,814
	R	1.663	1.381	0.024	0.144	28,776
	TM	1.910	1.524	0.011	0.140	57,632

Note: Chetty et al. (2014a) shrinkage applied.

Table 4: The Effects of a Change in Peer Value-Added on the Change in Own Value-Added

	All school-grade-year cells			Constant number of teachers		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	d(avg. resids)	d(avg. resids)	d(avg. resids)	d(avg. resids)	d(avg. resids)	d(avg. resids)
d(peer VA)	0.116***	0.092***	0.100***	0.124***	0.099***	0.122***
	(0.009)	(0.010)	(0.014)	(0.014)	(0.018)	(0.035)
d(num peers)	0.006***	0.011***	0.012***			
	(0.001)	(0.001)	(0.002)			
d(avg. lagged math)		-0.213***	-0.054***		-0.206***	-0.023
		(0.008)	(0.010)		(0.011)	(0.020)
d(avg. lagged reading)		0.073***	0.006		0.065***	-0.006
		(0.008)	(0.011)		(0.012)	(0.021)
d(frac nonwhite)		-0.099***	-0.008		-0.088***	0.017
		(0.016)	(0.020)		(0.023)	(0.043)
d(log num studs)		-0.088***	-0.057***		-0.117***	-0.078***
		(0.006)	(0.008)		(0.010)	(0.018)
Constant	0.003***	0.003***	0.003***	-0.001	0.000	-0.000
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
Observations	93,629	86,238	93,629	45,736	41,987	45,736
R-squared	0.003	0.074	0.386	0.003	0.127	0.501
principal effects	no	yes	no	no	yes	no
school-grade effects	no	no	yes	no	no	yes
school-year effects	no	no	yes	no	no	yes

Note: Standard errors in parentheses are clustered at the school level. Regressions weighted by average class size. Dependent variable in each column is the change in average classroom residuals from student-level OLS VAM specification. Independent variable of interest is the unweighted difference in shrunken VA between newcomers and leavers. Chetty et al. (2014a) shrinkage applied.

Table 5: The Effects of a Change in Peer Value-Added on the Change in Own Value-Added - by classroom assignment policy

	All school-grade-year cells			Constant number of teachers		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	d(avg. resids)	d(avg. resids)	d(avg. resids)	d(avg. resids)	d(avg. resids)	d(avg. resids)
d(peer VA)	0.101***	0.072***	0.082***	0.089***	0.054*	0.036
	(0.018)	(0.020)	(0.026)	(0.027)	(0.031)	(0.064)
T	0.002	0.007		0.003	0.020	
	(0.003)	(0.008)		(0.005)	(0.015)	
TM	0.001	0.003		0.002	0.005	
	(0.001)	(0.005)		(0.002)	(0.009)	
T*d(peer VA)	0.007	-0.001	0.117*	0.078	0.095	0.201
	(0.045)	(0.047)	(0.067)	(0.066)	(0.077)	(0.150)
TM*d(peer VA)	0.021	0.029	0.019	0.048	0.062	0.120
	(0.021)	(0.024)	(0.032)	(0.032)	(0.038)	(0.078)
d(num peers)	0.006***	0.011***	0.012***			
	(0.001)	(0.001)	(0.002)			
d(avg lagged math)		-0.213***	-0.054***		-0.208***	-0.025
		(0.008)	(0.010)		(0.012)	(0.021)
d(avg lagged reading)		0.075***	0.005		0.069***	-0.006
		(0.008)	(0.011)		(0.012)	(0.021)
d(frac nonwhite)		-0.098***	-0.006		-0.087***	0.017
		(0.016)	(0.020)		(0.024)	(0.043)
d(log num studs)		-0.089***	-0.058***		-0.118***	-0.083***
		(0.006)	(0.008)		(0.010)	(0.018)
Constant	0.002**	0.001	0.004***	-0.002	-0.004	-0.003***
	(0.001)	(0.003)	(0.000)	(0.002)	(0.006)	(0.000)
Observations	92,814	85,509	92,814	45,010	41,332	45,010
R-squared	0.003	0.073	0.381	0.003	0.126	0.494
principal effects	no	yes	no	no	yes	no
school-grade effects	no	no	yes	no	no	yes
school-year effects	no	no	yes	no	no	yes

Note: Standard errors in parentheses are clustered at the school level. Regressions weighted by average class size. Dependent variable in each column is the change in average classroom residuals from student-level OLS VAM specification. Independent variable of interest is the unweighted difference in shrunken VA between newcomers and leavers, and its interactions with assignment practice dummies (whether the school is tracking-only (T) or tracking & matching (TM); reference group is random (R) schools). Sample: Teachers in school-grade-year cells, where the number of teachers is constant. Chetty et al. (2014a) shrinkage applied.

Table 6: The Effects of a Change in Peer Value-Added on the Change in the Log of Own Class Size - by classroom assignment policy

VARIABLES	(1) d(log own class size)	(2) d(log own class size)	(3) d(log own class size)
d(peer VA)	-0.009	-0.015	-0.005
	(0.027)	(0.024)	(0.058)
T	-0.001	0.022	
	(0.002)	(0.010)	
TM	0.000	0.002	
	(0.001)	(0.009)	
T*d(peerVA)	-0.047	-0.008	0.022
	(0.042)	(0.040)	(0.083)
TM*d(peerVA)	-0.019	0.019	0.017
	(0.033)	(0.031)	(0.073)
d(avg. lagged math)		0.013	-0.002
		(0.010)	(0.024)
d(avg. lagged reading)		0.046***	0.040*
		(0.011)	(0.024)
d(frac nonwhite)		-0.052**	-0.082*
		(0.022)	(0.044)
d(log num studs)		0.610***	0.700***
		(0.012)	(0.025)
Constant	0.011***	0.004	0.009***
	(0.002)	(0.006)	(0.000)
Observations	45,010	41,332	45,010
R-squared	0.000	0.237	0.527
principal effects	no	yes	no
school-grade effects	no	no	yes
school-year effects	no	no	yes

Note: Standard errors in parentheses are clustered at the school level. Regressions weighted by average class size. Dependent variable in each column is the change in log class size. Independent variable of interest is the unweighted difference in shrunken VA between newcomers and leavers, and its interactions with assignment practice dummies (whether the school is tracking-only (T) or tracking & matching (TM); reference group is random (R) schools). Sample: Teachers in school-grade-year cells, where the number of teachers is constant. Chetty et al. (2014a) shrinkage applied.

Table 7: The Effects of a Change in Peer Value-Added on the Change in the Fraction of Nonwhite Students in Own Classroom - by classroom assignment policy

VARIABLES	(1) d(own class frac nonwhite)	(2) d(own class frac nonwhite)	(3) d(own class frac nonwhite)
d(peer VA)	0.012	0.015*	0.013
	(0.012)	(0.009)	(0.015)
T	0.002	0.000	
	(0.004)	(0.003)	
TM	-0.002	-0.002	
	(0.002)	(0.002)	
T*d(peerVA)	0.009	-0.002	0.003
	(0.018)	(0.014)	(0.029)
TM*d(peerVA)	-0.011	-0.013	-0.015
	(0.014)	(0.011)	(0.023)
d(avg. lagged math)		-0.004	0.003
		(0.003)	(0.007)
d(avg. lagged reading)		-0.001	0.001
		(0.003)	(0.008)
d(frac nonwhite)		0.941***	0.973***
		(0.008)	(0.016)
d(log num studs)		-0.010***	-0.003
		(0.003)	(0.007)
Constant	0.009***	0.001	-0.000
	(0.001)	(0.002)	(0.000)
Observations	45,010	41,332	45,010
R-squared	0.000	0.322	0.436
principal effects	no	yes	no
school-grade effects	no	no	yes
school-year effects	no	no	yes

Note: Standard errors in parentheses are clustered at the school level. Regressions weighted by average class size. Dependent variable in each column is the change in the fraction of nonwhite students in the class. Independent variable of interest is the unweighted difference in shrunken VA between newcomers and leavers, and its interactions with assignment practice dummies (whether the school is tracking-only (T) or tracking & matching (TM); reference group is random (R) schools). Sample: Teachers in school-grade-year cells, where the number of teachers is constant. Chetty et al. (2014a) shrinkage applied.

Table 8: The Effects of a Change in Peer Value-Added on the Change in Prior Math and Reading Achievement in Own Classroom - by classroom assignment policy

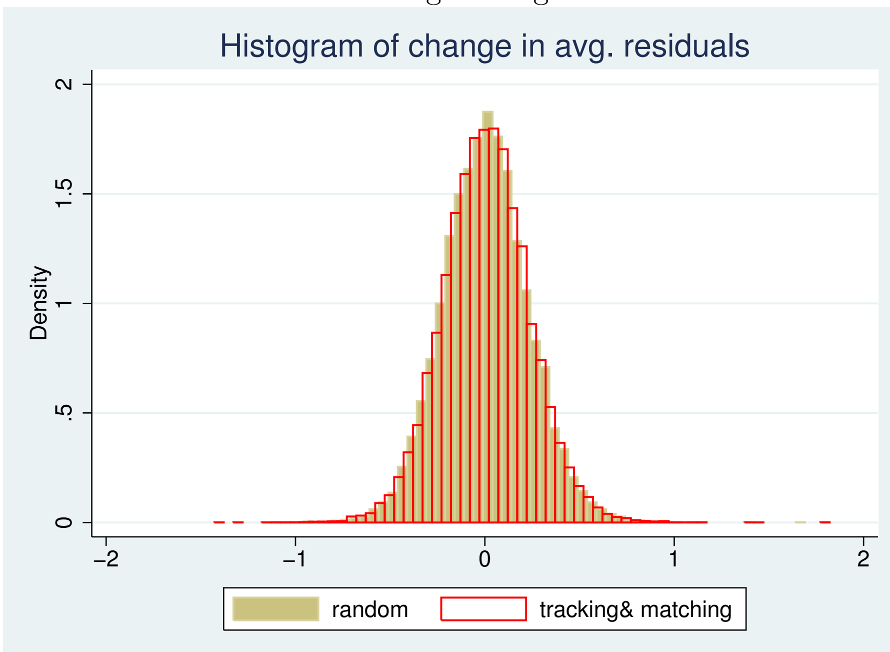
	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	d(own class lagged math)	d(own class lagged math)	d(own class lagged math)	d(own class lagged reading)	d(own class lagged reading)	d(own class lagged reading)
d(peer VA)	-0.001	0.006	-0.008	0.017	0.008	0.001
	(0.034)	(0.023)	(0.023)	(0.028)	(0.019)	(0.042)
T	-0.001	0.000		0.000	0.014	
	(0.007)	(0.012)		(0.005)	(0.010)	
TM	0.005	-0.002		0.001	-0.009*	
	(0.003)	(0.006)		(0.003)	(0.005)	
T*d(peerVA)	-0.016	-0.057	-0.067	-0.132**	-0.097**	-0.008
	(0.106)	(0.058)	(0.058)	(0.066)	(0.048)	(0.109)
TM*d(peerVA)	-0.036	-0.061**	-0.061**	-0.057	-0.044	0.009
	(0.042)	(0.030)	(0.030)	(0.037)	(0.027)	(0.062)
d(avg. lagged math)		0.960***	0.954***		-0.029***	-0.057***
		(0.007)	(0.020)		(0.008)	(0.021)
d(avg. lagged reading)		-0.034***	-0.041**		0.947***	0.968***
		(0.008)	(0.020)		(0.009)	(0.021)
d(frac nonwhite)		0.027	0.009		0.022	-0.024
		(0.017)	(0.041)		(0.020)	(0.041)
d(log num studs)		0.023***	-0.022		0.022**	-0.012
		(0.007)	(0.018)		(0.009)	(0.018)
Constant	-0.002	0.006	0.004***	-0.001	0.008**	0.004***
	(0.003)	(0.004)	(0.000)	(0.002)	(0.003)	(0.000)
Observations	45,010	41,332	45,010	45,010	41,332	45,010
R-squared	0.000	0.344	0.336	0.000	0.304	0.420
principal effects	no	yes	no	no	yes	no
school-grade effects	no	no	yes	no	no	yes
school-year effects	no	no	yes	no	no	yes

Note: Standard errors in parentheses are clustered at the school level. Regressions weighted by average class size. Dependent variable in each column is the change in the classroom average of prior math scores. Independent variable of interest is the unweighted difference in shrunken VA between newcomers and leavers, and its interactions with assignment practice dummies (whether the school is tracking-only (T) or tracking & matching (TM); reference group is random (R) schools). Sample: Teachers in school-grade-year cells, where the number of teachers is constant. Chetty et al. (2014a) shrinkage applied.

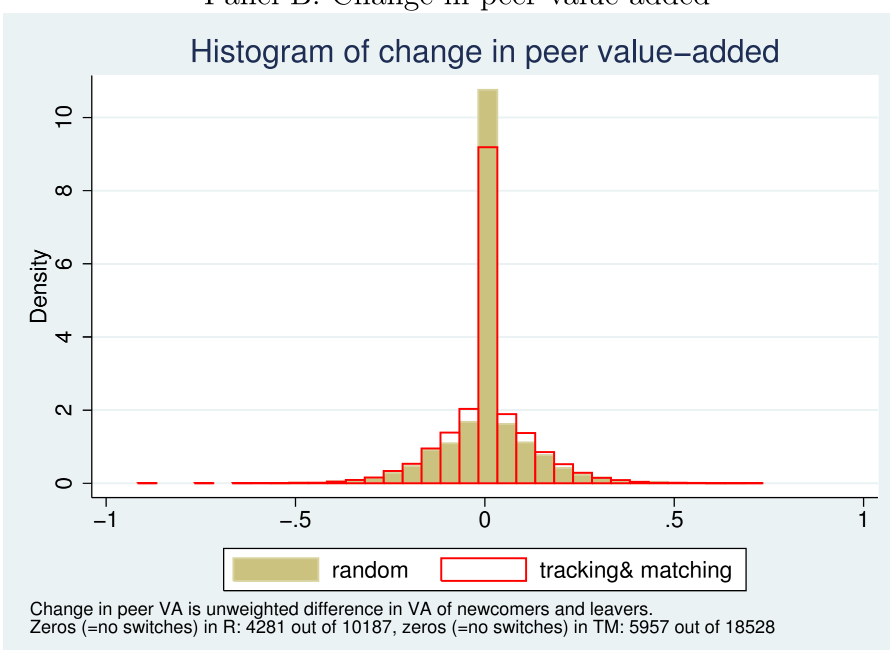
Figures

Figure 1: Histogram of changes in own- and peer value-added

Panel A: Change in avg. residuals



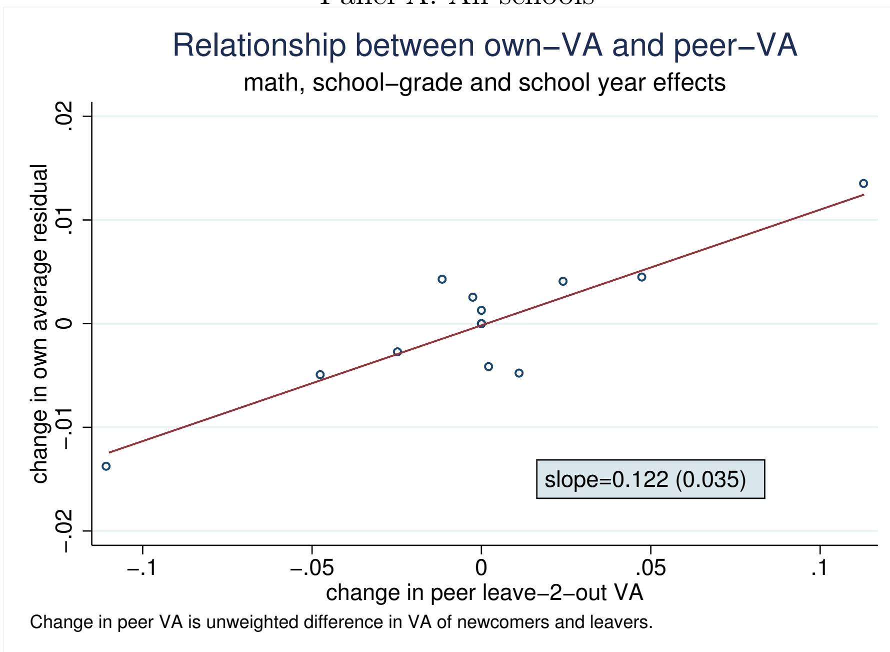
Panel B: Change in peer value-added



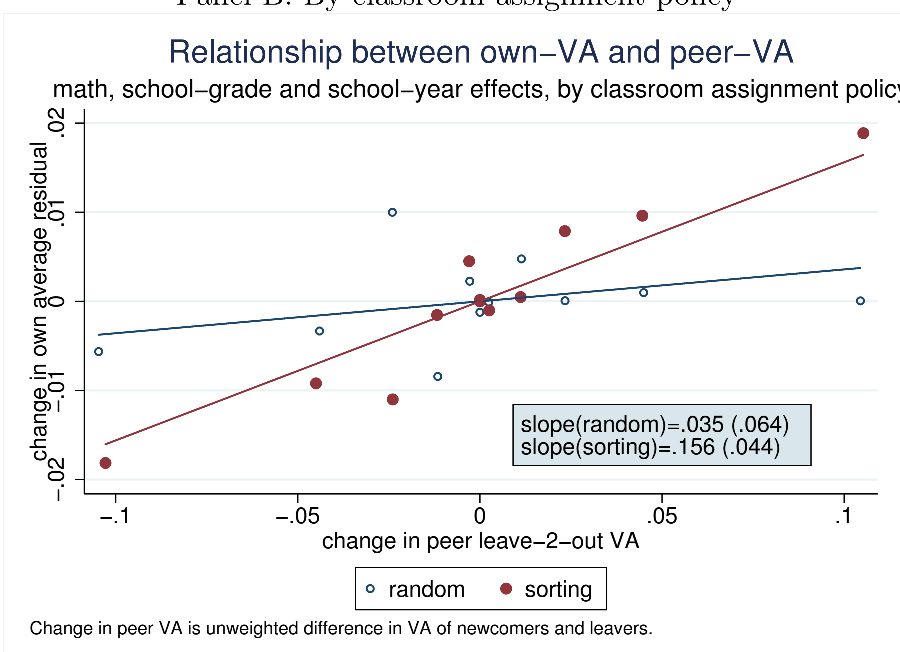
Notes: Panel A plots the histogram of the change in average classroom residuals from student-level OLS VAM specification. Panel B plots the histogram of the unweighted difference in shrunken VA between newcomers and leavers. Chetty et al. (2014a) shrinkage applied. Spike around 0 in Panel B is due to school-grade-years where there was no change in the teacher body.

Figure 2: Relationship between changes in own- and peer value-added

Panel A: All schools



Panel B: By classroom assignment policy



Note: Binned scatter graphs relating the change in average classroom residuals from student-level OLS VAM specification (vertical axis) and the unweighted difference in shrunken VA between newcomers and leavers (horizontal axis). Chetty et al. (2014a) shrinkage applied. Variables are residualized before plotting by the change in school-grade-year average prior math and reading achievement, change in the fraction of nonwhite students in the school-grade-year cell, change in the log number of students at the school-grade-year cell, and principal effects. Only school-grade-year cells with a constant number of teachers are included in the sample.