Cornell University
ILR School

Cornell University ILR School
DigitalCommons@ILR

Working Papers

ILR Collection

2-15-2009

Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers

Clement (Kirabo) Jackson
Cornell University, kirabo-jackson@northwestern.edu

Elias Bruegmann
Harvard University

Follow this and additional works at: http://digitalcommons.ilr.cornell.edu/workingpapers

Part of the Teacher Education and Professional Development Commons

DigitalCommons@ILR is celebrating its 10th anniversary!
Please share your DigitalCommons@ILR story!

This Article is brought to you for free and open access by the ILR Collection at DigitalCommons@ILR. It has been accepted for inclusion in Working Papers by an authorized administrator of DigitalCommons@ILR. For more information, please contact hlmdigital@cornell.edu.

Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers

Abstract
Using longitudinal elementary school teacher and student data, we document that students have larger test score gains when their teachers experience improvements in the observable characteristics of their colleagues. Using within-school and within-teacher variation, we further show that a teacher's students have larger achievement gains in math and reading when she has more effective colleagues (based on estimated value-added from an out-of-sample preperiod). Spillovers are strongest for less-experienced teachers and persist over time, teachers perform best when they are the weakest of their peer group, and historical peer quality explains away about twenty percent of the own-teacher effect. These results suggest peer learning.

Keywords
academic performance, teacher performance, peer quality

Disciplines
Education | Teacher Education and Professional Development

Comments

Suggested Citation
Jackson, C. K. & Bruegmann, E. (2009). *Teaching students and teaching each other: The importance of peer learning for teachers*. Retrieved [insert date], from Cornell University, School of Industrial and Labor Relations site: http://digitalcommons.ilr.cornell.edu/workingpapers/77/

This article is available at DigitalCommons@ILR: http://digitalcommons.ilr.cornell.edu/workingpapers/77

TEACHING STUDENTS AND TEACHING EACH OTHER: THE IMPORTANCE OF PEER LEARNING FOR TEACHERS[Ψ]

C. Kirabo Jackson
Cornell University
ckj5@cornell.edu

Elias Bruegmann
Cornerstone Research
ebruegmann@cornerstone.com

Draft: 2/15/2009

Abstract: Using longitudinal elementary school teacher and student data, we document that students have larger test score gains when their teachers experience improvements in the observable characteristics of their colleagues. Using within-school and within-teacher variation, we further show that a teacher's students have larger achievement gains in math and reading when she has more effective colleagues (based on estimated value-added from an out-of-sample pre-period). Spillovers are strongest for less-experienced teachers and persist over time, teachers perform best when they are the weakest of their peer group, and historical peer quality explains away about twenty percent of the own-teacher effect. These results suggest peer learning.

I. Introduction

Economists have long been concerned with human capital spillovers, given that these have strong implications for the optimal distribution of workers both within and across firms. When workers and their colleagues are complementary inputs in production, improvements in coworker quality may increase one's own productivity. There is evidence of such spillovers. Workers' wages are higher in firms with more educated coworkers [Battu, Belfield, and Sloane, (2003)], and wages for educated workers are higher in cities where the share of educated workers is higher [Moretti (2004b)]. Using direct measures of productivity, Azoulay, Wang and Zivin (2007) find that scientists have fewer grants and publications after a high-profile scientist leaves their institution. Peer quality may affect worker productivity, even if worker output is independent, by changing the social context. It has been documented that supermarket checkout workers worked faster while in the line of sight of a high-productivity worker [Mas and Moretti (2009)], the productivity of berry pickers converges to the productivity of their close friends when those friends are present [Bandiera, Barankay, and Rasul (2007)], and the shirking of workers who move branches is positively correlated with the average shirking of their coworkers [Ichino and Maggi (2000)]. However, Guryan, Kroft, and Notowidigdo (2007) find no evidence

[Ψ] We thank David Cutler, Li Han, Larry Katz, Andrew Oswald, Gauri Kartini Shastry, Kate Emans Sims, Daniel Tortorice, and participants in the Harvard Labor and Organizational Economics Lunches. Elias gratefully acknowledges financial support from the Graduate Society Dissertation Completion Fellowship of Harvard University. **Note:** In the interest of fairness, the ordering of the authors' names was determined by a coin flip.

of peer effects between randomly assigned golf partners in professional tournaments, suggesting the importance of context.

Although much research on empirical peer effects has focused on motivation and shirking, peer learning is an important mechanism. According to modern macroeconomic growth models [Lucas (1988); Romer (1990)], knowledgeable and skilled individuals increase the skill and knowledge of those with whom they interact, generating more ideas and faster macroeconomic growth. Despite the importance of peer knowledge spillovers both for firms' personnel practices and for the economy as a whole, there is little documented evidence of their existence.[1] Documenting peer learning is difficult because (1) output may be produced jointly, (2) there may be self-selection such that observed peer ability may be endogenous to unobserved ability, (3) peer knowledge is difficult to observe, and (4) unobserved factors could affect both output and peer quality.

We fill this gap in the literature, providing evidence of peer learning among teachers, using a unique longitudinal dataset of student test scores linked to teacher characteristics in North Carolina. Specifically, we test whether changes in a teacher's peers affect the test score growth of her own students, and we investigate possible mechanisms. Our empirical strategy is to estimate a student achievement value-added model with the inclusion of teacher peer attributes as covariates. To avoid the reflection problem [Manski (1993)], we use two measures of peer quality that are not determined by contact with peers: (1) observable peer characteristics that change exogenously, such as experience and certification test scores, and (2) unobservable peer quality based on teacher-specific, time-invariant value-added estimates from pre-sample data. We ensure that spillovers are not driven by students having direct contact with their teacher's colleagues by focusing on elementary school students who only have one teacher for the entire year.[2] To ensure that we do not use changes in peer quality due to teacher self-selection, we identify the *changes* in the performance of a teacher's students that are correlated with *changes* in the composition of her peers *within the same school* by including teacher-school fixed effects. Lastly, we define a teacher's peers to be all other teachers at the same school with students in the

[1] There is evidence of learning-by-doing spillovers across firms in the same industry [Thornton and Thompson (2001); Zimmerman (1982); Irwin and Klenow (1994)].

[2] We also remove all classrooms with teacher aides or team teachers to further eliminate the possibility of direct contact between students and their own teacher's colleagues. Students with an Alternative Education Program may be exposed to guidance counselors and special educators. This is not a problem, however, because none of these other teachers are used in our data to form the peer group.

same grade. This allows us to deal with the possibility that changes in a teacher's peers' attributes could be correlated with changes in school attributes or school policies (i.e., a school decides to de-emphasize math and gets rid of its best math teacher), by including year fixed effects for each school. Because teachers may be affected by teachers in other grades, our narrow definition of peers will provide a *lower* bound on the estimate of the importance of peers. We identify the effect of peers by comparing the changes in a teacher's students' test scores over time as her peers (and therefore the characteristics of her peers) change within the same school, while controlling for school-specific time shocks.

Although we are careful to control for a variety of possible confounding influences, we do not have random assignment of students to teachers or of teachers to peers. Because the possibility of spurious correlation remains, we present several specification and falsification tests. These indicate that our results are not driven by (1) endogenous peer quality changes across grades within schools or (2) the non-random dynamic sorting of students into classrooms.

Policy makers and researchers widely accept that there is substantial variation in teachers' ability to improve student achievement. There is evidence that years of experience, the selectivity of the teacher's undergraduate institution, teachers' test scores, certification, and principals' subjective evaluations are associated with student achievement [Brewer and Ehrenberg (1994); Hanushek (1997); Brewer and Goldhaber (2000); Anthony and Goldhaber (2007); Clotfelter, Ladd and Vigdor (2006, 2007); Jacob and Lefgren (2008)]. Studies that identify individual teachers associated with student test-score gains find that a one standard deviation increase in teacher quality leads to about one tenth of a standard deviation increase in math and reading scores [Rockoff (2004); Arronson, Barrow and Sander (2007); Hanushek, Kain and Rivkin (2005)].

Even though most models of teacher value-added rule out the existence of spillovers across teachers, much education policy is predicated on the notion that teachers learn from their peers. Instructional supervision and peer coaching are commonly used in schools to help teachers become more effective [McQuarrie and Wood (1991); Wood and Killion (1998)]. If teachers learn from their peers, the common practice of comparing the test score gains of teachers at the same school is misguided, because both a teacher's students *and* the students of that teacher's peers may be affected by the given teacher's attributes. This paper contributes to the nascent

literature[3] questioning the validity of standard value-added models by evaluating the assumption of no spillovers across teachers—a key identification assumption in teacher value-added models.

We find that students perform better when their teachers' peers have better observable characteristics. In models that use teacher value-added (based on historical student achievement gains) as a measure of teacher quality, we find that students experience greater test score gains when their teacher's peers have higher mean estimated value-added in both math and reading. These effects are robust across a variety of specifications and to our two distinct measures of teacher peer quality. Despite the predictive power of a teacher's peers, a failure to account for *contemporaneous* peer quality has a negligible effect on the own-teacher effect.

To help disentangle peer learning from other forms of spillovers, we test for empirical predictions that are most consistent with peer learning. We find that (1) less experienced teachers who are still acquiring "on-the-job" skills are most sensitive to changes in peer quality, (2) teachers with greater labor-market attachment are more sensitive to peer quality, (3) both current and historical peer quality changes affect current student achievement, (4) teachers perform better when their colleagues have better estimated effectiveness than themselves, and (5) historical peer quality explains away between 18 and 25 percent of the own-teacher effect. These findings are consistent with either direct learning from peers or what we refer to as peer-induced learning (learning induced by one's peers influencing one's decision to acquire work-related skills). This paper provides some of the first credible empirical evidence of learning associated with one's peers in the workplace.

The finding of asymmetric peer effects has important implications for how teachers are assigned. If inexperienced teachers are more influenced by peers than are experienced teachers, it is important that new teachers be assigned to schools with good mentor teachers. Similarly, if good experienced teachers help their peers more than they are negatively affected by them, spreading these good teachers across schools and grades would improve teaching efficiency. The findings here should give pause to advocates of *individual-level* merit-based pay because such pay schemes could reduce teachers' incentives to help their colleagues and could undermine peer

[3] Rothstein (2007) finds that value-added models perform particularly poorly in the presence of student tracking, such that future teachers have as much predictive power as current teachers in many standard value-added models. In contrast, Kane and Staiger (2008) use data from a random-assignment experiment and find that several non-experimental specification estimates of teacher effectiveness have strong predictive power in an experimental setting where students are randomly assigned to teachers. They also find that patterns of fade-out over time are very similar across experimental and non-experimental settings. Koedel (2008) tests for joint production among secondary school teachers but finds no evidence of cross-subject spillovers among high school teachers.

learning.

The remainder of the paper is organized as follows: Section II presents the theoretical framework, section III presents the identification strategy, section IV presents the data, section V presents our different measures of peer quality, section VI present the results, section VII presents specification and falsification tests, section VIII tests the predictions of the learning model and presents some extensions, and section IX concludes.

II. Theoretical Framework

We aim to observe how, and try to explain why, the performance of an individual teacher's students is affected by arguably exogenous changes in the quality of that teacher's peers.[4] In this section, we outline three potentially important sources of spillovers between teachers and outline a framework for thinking about learning between teachers.

1. ***Joint production and shared resources.*** Even when teachers have direct contact only with their own students, they may affect the time and other resources available to their peers' students. Teachers may share duties outside the classroom that require time and effort, so better peers may reduce the burden of these shared tasks. Similarly, the resources that teachers get from the school may be affected by the activities of their colleagues. The direction of this effect is ambiguous because more effective teachers may be better at lobbying for shared resources, increasing the amount available for each teacher, or may take a greater share of the resources available to the grade. A joint production explanation should yield a very simple prediction that a teacher may be positively or negatively affected by the quality of her contemporaneous peers. Under such an explanation, there may be substantial response heterogeneity, reflecting the fact that particular types of teachers are likely to be given certain types of tasks. Another prediction is that all peer effects should be contemporaneous, such that they do not persist over time.

2. ***Motivation and Effort.*** A teacher's peers can also affect her classroom performance by changing her own teaching effort. The presence of good teachers may motivate their colleagues through contagious enthusiasm or through embarrassment over the unfavorable direct performance comparison. Because overall school or grade performance may be used to evaluate

[4] There is a large literature on peer effects for students [this includes Angrist and Lang (2004); Hoxby (2000); Hoxby and Weingarth (2005); Sacerdote (2001); Lavy and Schlosser (2007); Zimmerman, (2003)]. There is also a literature documenting the importance of social networks [this includes Costa and Kahn (2007); Duflo and Saez (2003); Sorensen (2006); Laschever (2005)].

schools, the introduction of a better teacher to the grade could make free-riding more attractive. However, Kandel and Lazear (1992) suggest that peer pressure may force teachers to internalize their spillovers. If peer pressure is sufficiently strong, it could push teachers with better peers toward higher performance.[5] A motivation or effort explanation will have ambiguous empirical predictions; however, the empirical work on such mechanisms in the workplace suggests that teachers perform better if they have better peers. A simple motivation or effort explanation implies that all peer effects should be contemporaneous.

3. ***Peer Learning.*** Improvement in teacher effectiveness over time, particularly in the first few years of teaching, is a consistent finding in the literature. This finding suggests that on-the-job learning is very important for teachers. Therefore, we are interested in whether learning is a major avenue for the transmission of peer effects. We believe that learning has several important features that help distinguish it from other peer-effects explanations, and we examine these empirically. (1) On average, one can learn more from better peers, so we should observe positive correlation between peer quality and own-student performance. (2) Learning requires investment, so teachers with greater labor-force attachment and less experience (who have more years of teaching remaining and therefore have more years in which to benefit from investing in their teaching skills and learning from their peers) should be more likely to invest in learning and more sensitive to peer quality. (3) Learning is cumulative, so students should be affected by the composition of their teacher's past peers. (4) Because teaching ability is a combination of innate ability and learned skills, historical peer quality should explain some of the own-teacher effect. Lastly, (5) because teachers likely learn more from teachers who know more than them, the marginal effect of peer quality should be smaller when one is the best of one's peers and largest when one is the worst among one's peers.

One can easily distinguish a simple motivation story or a simple shared-task story from a learning explanation by testing empirical features (1) through (5) above. However, although these patterns imply a learning explanation, they do not necessarily imply learning directly from one's peers. It is possible that having better peers allows teachers to spend less time on other shared tasks and more time learning how to be a better teacher. Also, it is possible that when teachers have good peers, they are motivated to be better teachers and therefore invest in learning how to be a better teacher. Both these explanations are learning stories, but they explain peer-

[5] As found by Mas and Moretti (2009) and Bandiera et al. (2007) in other settings.

induced learning rather than direct learning from peers. Because understanding how teachers acquire human capital is important and relatively understudied, being able to distinguish any kind of learning from other explanations is useful. Because all learning explanations could yield the same empirical predictions, we are unable to distinguish a peer learning story from the peer-induced learning story. We are, however, able to test for peer-related learning (either through peers inducing a teacher to learn or through peers teaching their peers).

III Identification Strategy

In our analysis, a teacher's peers are defined as those teachers in the same school who teach students in the same grade in the same year. As discussed below, excluding peers from other grades is crucial to our identification strategy because that allows us to control for school-specific time shocks that could affect both student outcomes and teacher peer quality. Using variation in the quality of *all* a teacher's potential peers (teachers in all grades in the school) could lead one to confound school shocks with changes in peer quality. This clearly is undesirable. Teachers are more likely to be affected by their peers in the same grade, but because teachers may be affected by teachers in other grades, our estimates, using own-grade teachers, will provide a *lower* bound of the full effect of peers. Because establishing that peer effects exist is of first-order importance, and quantifying the full effect is secondary, we focus only on that variation that is credibly exogenous to other changes (i.e., variation in own-grade peer quality conditional on school-specific shocks). Our main specification uses the mean of peer characteristics as a useful summary of peer quality, but section VIII explores the use of other statistics.

To infer the effect of a teacher's peers on student test scores, we begin with a value-added model augmented to include measures of teacher peer quality. We start with our naïve baseline specification.

$$A_{it} = \delta A_{it-1} + \phi X_{it} + \eta W_{jt} + \mu \overline{P}_{j't} + \xi_{gt} + \varepsilon_{ijgst} \qquad [1]$$

In [1], we simplify the notation so that A_{it} represents A_{ijgst}, which is the achievement score of student i with teacher j in grade g of school s in year t. Similarly, A_{it-1} represents $A_{ij_{t-1}g_{t-1}s_{t-1}t-1}$, which is the score of student i with teacher j_{t-1} in grade g_{t-1} of school s_{t-1} in the previous year. X_{it} is a vector of student characteristics such as ethnicity, gender, and parental education level, W_{jt}

is a vector of characteristics of teacher j in year t, ξ_{gt} is a grade-by-year fixed effect, and ε_{ijgst} is the idiosyncratic error term. $\overline{P}_{j't}$ is a measure of peer quality. We discuss our measures of peer quality in Section V.

One of the major problems in identifying credible peer effect estimates is the fact that individuals often self-select to their peers. To avoid bias due to self-selection to peers, we remove all potentially problematic variation in peer characteristics that occurs as a result of the teacher's own movement. Specifically, we identify our effects based on *changes* in the characteristics of a teacher's peers and *changes* in the performance of her students, when the teacher has remained at the same school over time. This is accomplished by adding a teacher-school fixed effect to [1]. By relying only on variation within the scores of students of a given teacher within a given school, all variation in peer quality comes from either (1) a teacher being re-assigned to another grade within the same school or (2) the movement of peers in or out of her school and grade.

Another major difficulty in identifying peer effects, particularly where individuals are not randomly assigned to peers, is that changes in peer quality may be correlated with omitted factors that also affect own outcomes. For example, improvements in peer quality within a school may coincide with improvements in other dimensions in the school. For example, a disruptive event, such as a hurricane, could cause good teachers to leave the school at the same time that students perform poorly. Any school-specific shock that has a deleterious effect on both peer quality and student achievement would produce results that look like positive peer effects. To address this concern, we make comparisons only within groups of teachers at the same school at the same time (i.e., teachers who are subject to the same school-level shocks but teach in different grades and therefore have different peers). We do this by adding a school-by-year fixed effect to [1]. The school-by-year effect removes those confounding factors that affect all grades in the school that could also have an effect on teachers' peer quality. Because peer quality for each teacher in a particular school is identified at the school-grade-year level, we cannot include school-grade-year effects. With the inclusion of school-year and grade-year and school-grade fixed effects, our estimates will be biased only in the unlikely event that higher quality teachers are added to grades within schools at the same time as other improvements are made that are particular to that grade within the school. We present evidence that this is not the case in section VII.

Our preferred model is therefore an augmented version of the student achievement model in equation [1] that includes teacher peer quality as an input, while controlling for teacher-school fixed effects and school-by-year fixed effects. Specifically, we estimate [2] below.

$$A_{it} = \delta A_{it-1} + \phi X_{it} + \xi_{ts} + \mu \overline{P}_{j't} + \xi_{gy} + \xi_{sy} + \varepsilon_{ijgst} \qquad [2]$$

All variables are as before, ξ_{ts} is a teacher-school fixed effect, and ξ_{sy} is a year fixed effect for each school. Although it is tempting to include as many fixed effects as possible to remove confounding factors, such an approach often leads to weak identification, undermining the overall objective of identifying the parameter of interest [Anderson and Wells (2007)]. Although our preferred specification includes teacher-by-school fixed effects and school-by-year fixed effects, to show that our results are robust across a variety of empirical specifications, we report results from a series of regressions with the same basic specification described in [1] but with different sets of fixed effects.

IV Data

We use data on all third- through fifth-grade students in North Carolina between 1995 to 2006. Data come from the North Carolina Education Research Data Center.[6] Our student data include demographic characteristics, standardized test scores in math and reading, and codes allowing us to link the data to information about the schools the students attend and the teachers who are administering their tests. According to state regulation, the tests must be administered by a teacher, principal, or guidance counselor. Discussions with education officials in North Carolina suggest that tests are always administered by the students' own teachers when these teachers are present.[7] Also, all students in the same grade take the exam at the same time; thus, any teacher teaching a given subject in a given grade will almost certainly be administering the exam only to her own students. This precludes our misspecifying a teacher as her own peer. We take several steps to limit our sample to teachers who we are confident are the actual teachers for the students being tested. We include only students who are being administered the exam by a teacher who teaches math and reading to students in that grade, and we remove teachers who are

[6] These data have been used by other researchers in different contexts to identify the effect of charter schools on student outcomes [Bifulco and Ladd. (2005, forthcoming)], to look at the effect of teachers on student outcomes [Clotfelter, Ladd, and Vigdor (2006, 2007); Clotfelter, Ladd, Vigdor, and Wheeler (2007); Rothstein (2007)], the effect of schools on student achievement [Hastings, Van Weelden, and Weinstein. (2007); Hastings and Weinstein. (2007)], the effect of student demographics on teacher quality [Jackson (2008)], and the effect of schools on housing prices [Kane, Staiger, and Riegg (2005)].

[7] Conversations with several school principals in the state reveal that this is true.

co-teaching or have a teaching aide. This process gives us roughly 1.37 million student-year observations matched to teachers we are confident taught the students the material being tested. We use *changes* in student test scores as the dependent variable, so our regression analysis is based on the outcomes of fourth and fifth graders.[8]

The students are roughly 62 percent white and 29.5 percent black, and they are evenly divided between boys and girls (similar to the full state sample). About 65 percent of students are the same race as their teacher, and about 50 percent are the same sex. The average class size is 23, with a standard deviation of 4. About 11 percent of the students' parents did not finish high school, 43 percent have just a high school diploma, roughly 30 percent have some post-high school education but no four-year college degree, and roughly 14 percent of the students have parents who have a four-year college degree or graduate degree as their highest level of education. The test scores for reading and math have been standardized to have a mean of zero and unit variance, based on *all* students in that grade in that year. The average year-to-year test score growth is zero, with standard deviation of 0.583 for math and 0.613 for reading. Students in our sample attend a total of 1,545 schools. Schools on average had 101 students and 6.6 teachers.

About 92 percent of teachers we successfully match to students are female, 83 percent are white, and 15 percent are black. Summary statistics for teacher-qualifications variables in our data are found in Table 1 at the teacher-year level. The average teacher in our data has thirteen years of experience, and roughly 6 percent of the teachers have no experience.[9] Roughly 20 percent of teachers have advanced degrees. The variable "regular licensure" refers to whether the teacher has received a regular state license or instead is working under a provisional, temporary, emergency, or lateral entry license. About 67 percent of the teachers in our sample have regular licensure. We normalize scores on the Elementary Education or the Early Childhood Education tests that all North Carolina elementary school teachers are required to take, so that these scores have a mean of zero and unit variance for each year in the data. Teachers in our sample perform near the mean, with a standard deviation of 0.81. Lastly, about 4 percent of teachers have

[8] There is a pre-test given to student before third grade; however, these data are not available for all years.
[9] Teacher experience is based on the amount of experience credited to the teacher for the purposes of determining salary; therefore, it should reflect total teaching experience in any school district.

National Board Certification.[10]

For part of our analysis, we use the average value of the characteristics of the other teachers in the same school and grade to indicate peer quality. Table 1 includes summary statistics for these measures. Often, the variation in our data will come from the movement of peers into or out of a school grade, so we look at several summary statistics to get a better understanding of this process in our data. First, we consider the distribution of peer group size. The mean teacher in our data has about three other teachers in the same school grade and year that appear in our data. More than 90 percent of teachers have six or fewer colleagues in our data. The small number of teachers per school grade suggests that the relevant quality of peers in a teacher's own grade may change substantially with the introduction or exit of just one or two good or bad teachers. As shown in Table 2, during the years 2001 through 2006, 65.8 percent of the teachers are in the same school and grade as the most recent previous year they appear in the data (going as far back as 1996), 6.0 percent are in the same school but teaching a different grade, 7.4 percent have moved from another school in our data since the most recent previous year, and 20.9 percent do not appear previously in our data. These high levels of mobility aid our empirical identification.

We are also interested in which teachers are moving between grades and schools. In Panel A of Table 2, we compare summary statistics for the observable teacher qualifications of each group of teachers: whether they are in the same grade and school as the previous year, the same school but different grade as the previous year, a different school as the previous year, or new to the data. The means for each of these variables are quite similar across groups (with the obvious exception of experience for teachers new to the data). This suggests that teachers who move between schools or grades are similar to teachers who do not. Even if the moving and non-moving teachers look similar in the data, however, the moving teachers could be moving to schools and grades with systematically better or worse peers. To test this hypothesis, we present (in Panel B of Table 2) the difference between each teacher's own characteristics and the average of her new peers' characteristics and the standard error of this difference. Teachers who move from a different grade in the same school differ from their peers only in that they are slightly more likely to have regular licensure. The absolute difference is quite small. Teachers moving

[10] Teachers in North Carolina are given a 12 percent pay raise if they achieve certification by the National Board for Professional Teaching Standards. This certification requires significant work on the part of the teacher and potentially signals teacher quality.

between schools are more likely to have advanced degrees and regular licensure than the peers in the grade and school they enter, but the differences are small. These statistics suggest that teachers who change schools or grades are similar to their new colleagues.

V Measures of Teacher Peer Quality

A naïve empirical strategy to test whether teachers exert spillover effects on each other's students would be to estimate standard student value-added regressions with the inclusion of the mean test score growth of a teacher's peers' students. We do not pursue this strategy because the performance of a teacher's peers' students is itself a function of the teacher's own attributes. We address this problem with two different measures of peer quality that are not co-determined with a teacher's own performance. The first approach is to use the observable characteristics of peer teachers, and the second is to use the value-added of peer teachers estimated in an *out-of-sample* pre-period. The two different approaches complement and provide a robustness check on each other. In both approaches, our model identifies the social interaction effect, which is a combination of the effect of group characteristics on individual outcomes and the effect of group behavior on individual behavior [Manski (1993)].

Observable characteristics as a measure of quality

For the first proxy for peer quality, we compute the average characteristics for each teacher's peers. For each school-year-grade cell, we compute the mean attributes of all other teachers in that cell, so that peer quality for teacher j in grade g at school s in year t, $\overline{W}_{j'gst}$, is the mean characteristic of all *other* teachers j' in grade g at school s in year t. These peer averages are summarized in Table 1. We include these peer averages as a measure of peer quality $\overline{P}_{j't}$ in equation [2]. Changes in this measure of peer quality occur when the characteristics of a teacher's peers change (e.g., becoming more experienced or obtaining regular licensure) or when the identity of a teacher's peers change. Because observable teacher characteristics such as experience vary exogenously with time, and because teachers are unlikely to obtain certification as a result of their peers, this approach is unlikely to be subject to the reflection problem. Our second measure of peer quality, however, relies solely on changes in the identity of a teacher's peers. The first approach has the advantage of being straightforward and allowing us to include data on almost all teachers, but, as in previous research, we find these characteristics are weak predictors of teacher quality. For this reason, we prefer our second approach for most of our

analysis.

Estimated value-added as a measure of quality

Our main proxy for teacher peer quality is the historical estimated value-added of a teacher's peers. Because a teacher's value-added could be due to exposure to high-ability peers, it is important to identify variation in peer quality (as measured by value-added) that is not subject to spillover bias in the estimation equation. We address this problem by using *out-of-sample* estimates of teacher value-added based on data between 1995 and 2000, while estimating the effect of *changes* in estimated peer value-added on *changes* in own-student outcomes using data from 2001 through 2006. Using *changes* in peer quality addresses the concern that the *level* of a teacher's peer quality could have been affected by her own quality in the pre-sample period.

Using pre-sample (1995 through 2000) data, we estimate teacher value-added by estimating a student achievement model of the form [1] with the inclusion of indicator variables denoting if the student *i* is in class with teacher *j* (for each teacher). A detailed description of the value-added estimation, including the estimation equation and the results, is included in the Appendix. The coefficients on the teacher indicator variables, the θ_j's, are standardized and normalized and are used as measures of teacher quality in the estimation sample (2001 through 2006 data). As with the observable teacher characteristics, for each school-year-grade cell, we compute the mean value-added of all other teachers in that cell, so that peer quality for teacher *j* in grade *g* at school *s* in year *t*, $\overline{\theta}_{j'gst}$, is the mean estimated value-added of all *other* teachers *j'* in grade *g* at school *s* in year *t*. These estimated teacher value-added effects do not vary over time, so all of the variation in mean peer value-added comes from changes in the identity of a teacher's peers and, as such, is not subject to the reflection problem.

This value-added approach has the disadvantage that teachers who are not in-sample between 1996 and 2000 will have no estimated value-added. New teachers, teachers from out of state, and non-elementary school teachers therefore will have no estimated value-added in our estimation sample (2001 through 2006). Because we would like to include all teachers in our estimation sample and would like to use all of a teacher's peers, we use the full sample of teachers, and we assign the mean of the distribution to teachers with no estimated teacher value-added as well as including control variables for the proportion of a teacher's peers with no estimated value-added. The proportion of teachers in a teacher's peer group with no estimated value-added serves as a proxy for the characteristics of teachers with missing peers. To ensure

that our treatment of teachers with missing value-added is not driving our results, we (1) estimated models that include dummies for having missing peers, (2) estimated models that use imputed teacher value-added based on observable teacher characteristics for those teachers with missing teacher effects, (3) estimated models that include the number of new teachers to the grade in a given year, and (4) estimated models that used the mean only of those teachers with estimated value-added. Across all these models, the results are virtually unchanged.[11]

VI Results

We begin our empirical section by considering the effect of the average of teachers' peers' observable characteristics on teachers' own performance (i.e., estimating equation [2] while using observable peer characteristics as our measure of peer quality). Table 3 presents these results. We report the results for math and reading test scores in the left and right panels, respectively. We present results obtained with school fixed effects, student fixed effects, and including both teacher-by-school and school-by-year effects. The effects of own-teacher characteristics in columns 1 and 4 are reasonable for both math and reading. Students have higher test scores in both subjects when their own teacher has a regular teaching license, has higher scores on her license exam, is fully National Board certified, and has more years of experience. Having a teacher with no previous experience is particularly detrimental, and having a teacher with an advanced degree appears to be *negatively* correlated with test scores, conditional on the other covariates.

We now turn our attention to the effect of peer characteristics. Although Table 3 reports results for all specifications, we focus on the results for our preferred model in columns 3 and 6. In this specification, for both math and reading, the coefficients on all the peer experience categories are positive and statistically significant. Because the omitted variable is the share of peers with no years of experience, this indicates that having more peers with more than one year of teaching experience has a statistically significant positive effect. The differences between other experience categories are smaller and generally not statistically significant. Average peer licensure score and share of peers with advanced degrees have small and statistically insignificant coefficients for both math and reading. One can reject the joint hypothesis that

[11] Using the mean only for those teachers with estimated value-added results in peer effects that are 14 percent smaller in math and 4 percent smaller in reading. Because ignoring teachers without value-added introduces additional measurement error, a reduction in the estimated effect is expected. In practice, the reductions are small.

teacher peer characteristics are equal to zero at the 1 percent level for both math and reading. Looking to specific characteristics, one can reject the joint hypothesis that teacher peer experience are equal to zero at the 10 percent level for both math and reading. One cannot reject, however, the joint hypothesis that teacher peer characteristics, other than peer experience, are equal to zero at traditional levels for either math or reading.

To summarize the cumulative effect of differences in observed peer quality, we do two things. First, using the coefficients from the regression, we compute the value-added associated with a teacher's peers, and then we look at the standard deviation of value-added to get a sense of the importance of observable teacher peer characteristics. The standard deviation of this imputed peer-value-added is 0.006 for both reading and math. This implies that a one-standard-deviation improvement in mean peer quality would improve test scores by a relatively small 0.006 of a standard deviation. Second, as shown in Table 4, we show the estimated difference in students' test scores between otherwise identical students with better or worse teacher peer groups. Specifically, we compute the effect of having bad peers (all peers have a value of zero for experience and have exam score, advanced degree share, regular licensure share, and National Board certification share that are one standard deviation below the mean) relative to average peers (i.e., all peers have four to nine years experience, and other peer attributes are at the sample mean), and good peers (peers have ten to twenty-four years experience, and other attributes are one standard deviation above the mean) relative to average peers.[12] In our preferred specification, having a teacher with bad peers leads to test scores that are 0.025 standard deviations lower for math and 0.009 deviations lower for reading (the math effect is statistically significant). Having the good peer group does not increase math performance relative to the average peer group but does raise reading by a statistically significant 0.01 standard deviations. Together, these results suggest that a standard deviation improvement in observable peer characteristics is associated with one tenth of a standard deviation improvement in own students test score in both math and reading. These effects of this second exercises are larger than the previous exercise because a standard deviation increase in mean peer quality is a smaller change than improving the characteristic of all a teacher's peers by one standard deviation. As previously noted, observable teacher characteristics are relatively weak predictors of a teacher's own quality,

[12] In cases where a value one standard deviation below the mean would be less than zero for zero to one variables, we set the characteristics for the worst peer group to zero.

so these results are likely to be a lower bound on the true peer effects. We now examine the results that use pre-period value-added as a potentially more powerful indicator of peer quality.

Peer Value-Added Results

Table 5 shows the effect of a teacher's peers' estimated value-added (estimated out of sample using 1995 through 2000 data) in math and reading on her own students' math and reading test score growth (using data from 2001 through 2006). To ensure that the teacher value-added results are not driven simply by the observable teacher characteristics, we estimated models that included both estimated peer quality and observable peer characteristics. The coefficients on the observable peer characteristics are not statistically significant when estimated peer quality is included, and the inclusion of observable teacher peer characteristics has very little effect on the peer value-added estimates.[13] This suggests that the peer value-added estimates *are not* driven by any of the observable peer characteristics summarized in the previous section. Because observable teacher peer characteristics have little predictive power conditional on estimated teacher value-added, and because including them does not change the results in any meaningful way, we omit observable teacher peer characteristics from this part of the analysis. Note that all models include the full set of controls from Table 3. Columns 1 through 3 summarize the math peer effects, and columns 4 through 6 summarize the effects for reading. Column 1 shows the basic within-school math results. In this specification, a one-standard-deviation increase in a student's own-teacher's value-added increases her test scores by 0.1268 standard deviations, and a one-standard-deviation increase in a teacher's peers' mean value-added increases student scores by 0.0522 standard deviations. These estimated effects are likely to be underestimates because both teacher effectiveness and the effectiveness of a teacher's peers are measured with noise. The magnitude of the own-teacher effect is on the same order of magnitude as found in much of the literature. In this specification, the mean teacher peer effects are roughly equal to 40 percent of the own-teacher effect.

Column 2 includes student fixed effects. The statistically significant coefficient of 0.0604 on teacher peer value-added in column 3 shows that the teacher peer effects probably are not

[13] In models that include both teacher peer experience and teacher peer value-added, one cannot reject the null hypothesis of no teacher peer experience effects (conditional on peer value-added) at the 20 percent level for either math or reading. In contrast, for both math and reading, the hypothesis that peers' value-added is equal to zero (conditional on peer experience) is rejected at the 1 percent level.

driven by static student self-selection. In fact, the coefficient on the peer effect increases—a finding consistent with our contention that most plausible scenarios of student self-selection are likely to bias our estimates down rather than up. Because student tracking may be dynamic, such that a time-invariant student fixed effect would not control sufficiently for student self-selection, in section VII we present empirical evidence that our results are not confounded by dynamic tracking.

We now focus on the preferred specification in column 3. We remind the reader that the preferred model uses only within teacher and school variation to remove any selection of teachers to better peers, and it includes school-by-year effects to account for any school policies or school-specific shocks that could affect both peer quality and student test scores.[14] The coefficient on peer value-added for math in column 3 is 0.0398, suggesting that a one-standard-deviation increase in the mean estimated value-added of a teacher's peers is associated with a 3.98 percent of a standard deviation increase in math test scores. This is more than twice the size of using observable peer characteristics. For the average teacher with three peers, replacing one peer with another that has one standard deviation higher value-added will increase her students' math test scores by 1.3 percent of a standard deviation. This corresponds to between one tenth to one fifth of the own-teacher effect.

Columns 4 through 6 show the effects on reading test scores. The effects are qualitatively similar to those for math; however, the magnitudes are smaller (a consistent finding in the teacher quality literature). The basic within-school results in column 4 show that a one-standard-deviation increase in own-teacher value-added leads to a 0.0547 standard deviation increase in reading test scores and that a one standard deviation change in a teacher's peers' value-added leads to a 0.0262 standard deviation change in test scores. As with math, the student fixed-effects model in column 5 shows slightly larger and statistically significant peer effects. The preferred model, in column 6, includes teacher-by-school and school-by-year effects. It shows that a one-standard-deviation increase in mean peer value-added is associated with a statistically significant 0.026 standard deviation increase in student reading test scores. For the average teacher with three peers, replacing one peer with another that has one standard deviation higher value-added will increase her students' test scores by 0.86 percent of a standard deviation. As for math, this

[14] This was implemented using the "felsdvreg" command written by Thomass Cornellisen, described in Cornellisen (2006), based on the three-way error model proposed by Abowd, Creecy. and Kramarz (2002).

corresponds to between one tenth and one fifth of the own-teacher effect.[15]

One implication of significant teacher peer effects is that failing to take teacher peer inputs into account when estimating own-teacher value-added could lead to inconsistent estimates. Although peer effects are important in explaining variation in student test scores, the amount explained by teacher quality is virtually identical in models that include or do not include peer value-added. In math, the proportion of the variance in test score growth associated with the teacher fixed effects, $Cov(\Delta A_{ij,}\theta_j)/Var(\Delta A_{ij})$, is 0.141 when peer attributes are included and 0.1432 when they are not included. In reading, $Cov(\Delta A_{ij,}\theta_j)/Var(\Delta A_{ij})$ is 0.067 when peer attributes are included and 0.069 when they are not included. This suggests that the explanatory power of teacher effects is very slightly reduced when contemporaneous peer value-added is included, so that a student's own teacher accounts for about 14 percent of the variation in math test scores and just under 7 percent of the variation in reading test scores.[16]

VII Specifications and Falsification Tests

Because students are not randomly assigned to teachers and teachers are not randomly assigned to schools or classrooms, it is important that we isolate variation that is not confounded by student selection or teacher-self-selection, or correlated with confounding factors that also affect student achievement. Although including teacher-by-school effects credibly addresses the self-selection of teachers to peers, and although the inclusion of school-by-year effects credibly addresses the concern that schools that see improvements in peer quality may be improving in other areas, a few endogeneity concerns remain. We address these below.

Dynamic sorting could bias the estimated teacher effects

It is possible that our results could be confounded by dynamic tracking that would not be fully controlled for with a time-invariant student fixed effect or time-changing lagged test scores. Rothstein (2007) argues that in the presence of dynamic tracking, individual teacher fixed effects could be severely biased, and he presents compelling evidence of non-random student

[15] A model that includes student fixed effects and teacher-school effects yields a coefficient on math peers of 0.026 with a standard error of 0.008, along with a coefficient on reading peers of 0.0196 with a standard error of 0.01.

[16] If some students were taught by their homeroom teacher's peers but were wrongly classified as being taught by the homeroom teacher, the explanatory power of the own teacher would be lower when peer attributes are included. The fact that the explanatory power of the own teacher is unchanged when peers are also included is consistent with our contention that the spillovers are not due to the actual teacher being misclassified as the teacher's peer.

assignment to classrooms conditional on student fixed effects. For our purposes, we do not require that student assignment be random. We instead use the much weaker condition that the expected value of the unobserved student characteristics be uncorrelated with true teacher quality (i.e., $E[\varepsilon_i \mid \theta] = 0$). In the one paper that compares the predictive power of teacher value-added estimates obtained from random assignment to those obtained from non-random assignment, Kane and Staiger (2008) find little evidence of bias on average.

We present a falsification test of our identifying assumption ($E[\varepsilon_i \mid \theta] = 0$) in Table 6. Because a student's future teacher should have no causal effect on that student's current test score performance, any non-zero effect would indicate bias. We contend that if there is positive/negative selection, the estimated value-added of the teacher the student will have in the following year will be positively/negatively correlated with the student's achievement in the current year. If there is positive/negative selection, those teachers who are systematically associated with contemporaneous gains should be predictive of test score gains/losses for their *future* students.

Table 6 shows the coefficient of the estimated value-added (estimated out of sample) of a student's current teacher and her future teacher. The controls in this model include all student and teacher characteristics, school fixed effects, and grade-by-year fixed effects. (The main conclusions are invariant to the specification chosen.) Although the coefficient on the current teacher effect is 0.12 for math and 0.052 for reading (both significant at the 1 percent level), the coefficients for the *future* teacher's effect are only 0.002 and -0.002 for math and reading, respectively (both have p values greater than 0.3). In other words, teachers that we identify as effective have a strong positive effect on their current students' test scores but no effect on their future students' test scores. This result suggests that no systematic student sorting occurs. The null hypothesis of equality of the current teacher effect and the future teacher effect is rejected at the 1 percent level. In fact, the future teacher's coefficients are both less than one tenth the size of those for contemporaneous teachers, and the future teacher effects for reading and math have opposite signs. This is inconsistent with systematic selection because the math and reading teachers are in fact the same teacher. Furthermore, the standard errors on the future teacher value-added are small, indicating that the true values, if not actually zero, are very close to zero.

Dynamic sorting could produce patterns that look like positive peer effects

Dynamic sorting (or dynamic student selection) also could produce empirical patterns that look like positive contemporaneous peer effects. For example, consider the following scenario. Suppose that the principal assigns "difficult" students to teachers with the highest value-added and assigns the "easiest" students to less experienced or less able teachers. In such a scenario, when a strong older teacher retires and is replaced by a weaker and less experienced teacher (with lower value-added), incumbent teachers will be more likely to receive the "difficult" students. Sorting of students across classrooms in such a manner would make it look as though having weaker peers hurts the incumbent teacher if the econometrician is unable to control sufficiently for student ability.

This particular *dynamic* sorting story would be problematic because it would generate a *negative* correlation between true teacher quality and unobserved student ability. Specifically, in the sorting story described above, difficult students are assigned to those teachers who are most able, so that there is *negative* sorting. Although this negative sorting story is possible, most plausible sorting scenarios result in a *positive* correlation between teacher quality and unobservable student ability. First, affluent parents, who are more likely to have higher achieving children, are more likely to put pressure on principals to assign their children to "good" teachers. Second, those "good" teachers are able to request classrooms with easier-to-teach students, because these teachers are in higher demand and could threaten to leave. Clotfelter, Ladd, and Vigdor and Glennie (2008) document this pattern of *positive* selection *within schools* on observable teacher characteristics and observable student characteristics (e.g., low income, ethnic minority, and low-achieving students are placed in classrooms with less-qualified teachers). As such, both *a priori* notions of the nature of the bias and the empirical evidence suggest that insofar as there is any bias, it is likely to generate a *positive* correlation between teacher quality and student ability. Spurious correlation between teacher peer quality and student ability due to positive sorting would lead us to underestimate the contribution of the quality of a teacher's peers.

Fortunately, the test for bias caused by such dynamic sorting is the same as the test for no systematic bias in the estimated teacher effects. The results in Table 6 suggest that there is no such sorting. The positive peer effects we document therefore are unlikely to be driven by negative selection to teachers. Furthermore, in section VIII, we will show that a teacher's peers in the previous two years have a statistically significant and positive effect on her current

students. This is important because although negative selection and dynamic student sorting could potentially explain contemporaneous peer effects, it cannot explain persistent peer effects.

Changes in peer quality within schools may be endogenous, so that peer quality improvements coincide with other *grade specific* interventions

The remaining endogeneity concern is that schools may be more likely to shift good teachers across grades or to hire better new teachers into a grade (at the same time as the schools shift other resources) when particular grades are performing poorly relative to other grades in the school. In this scenario, even with controls for school-by-year effects, some peer effects could be confounded by other resources and efforts in that grade. We believe that bias resulting from new hiring being correlated with other grade-specific changes is unlikely because schools do not have much control over when teachers leave and because new hiring is likely to take place because of changes in class size or from vacancies occasioned by voluntary turnover. It is possible, however, that principals shift teachers across grades in response to poor grade performance at the same time that they implement other grade-specific improvements. We empirically test the possibility of endogenous peer changes both from outside the school and from within the school into a grade.

Table 7 shows the results of a regression at the school-grade-year level. We test for whether, conditional on school-by-year effects, a grade's performance in the previous year affects its likelihood of receiving a new teacher in the current year. If poor performance in the previous year was systematically correlated with an increased likelihood of having a new peer, one would worry that changes in peer quality are confounded with changes in resources and other efforts. In Table 7, the dependent variables are having a new peer in a grade from the same school, having a new peer in the grade from a different school in North Carolina, and having a new peer in the grade from outside the data. Included as independent variables are the proportion of teachers in various experience groups the previous year and the math and reading test score growth the previous year. Columns 4 through 6 also include the level of the math and reading test scores the previous year. As a further check, in Columns 7 through 9, we include the mean teacher value-added in the grade the previous year. Across all models, one cannot reject the null hypothesis that the arrival of a new teacher to a particular grade within a school is unrelated to the historical level and growth of test outcomes in both reading and math. Also, one cannot reject the null hypothesis that the arrival of a new teacher to a particular grade within a school is

unrelated to the estimated value-added of incumbent teachers in the grade. Teacher experience variables, however, do have predictive power, as one would expect, given that experience is a strong predictor of retirement.

VIII Extensions: Testing the Learning Hypothesis

We show in Section VIII that teachers perform better when their peers are better, in terms of both observed and unobserved quality. As discussed in section II, peer spillovers could exist for a variety of reasons. In this section, we test the five empirical predictions described in Section II that would be consistent with peer learning or peer-induced learning. All results from this point on are based on our preferred specification, with teacher-school and school-year fixed effects.

As discussed in section II, experienced teachers on average have fewer years of teaching remaining and therefore have fewer years in which to benefit from investing in their teaching skills and learning from their peers. Because learning from one's peers entails some costs, teachers with fewer years of experience (who tend to be earlier in their careers) will be more likely to invest in peer learning and should be more responsive to peer quality than more experienced teachers. This is a variation of the prediction that the incentives to invest in job-specific human capital decrease with tenure [Becker (1962); Jovanovic (1979)]. We test this hypothesis by interacting the own-teacher years of experience with the peer quality variables (while controlling for own experience). In Table 8, we report coefficients on the interactions, where the omitted group is teachers with no years of experience. The results are generally supportive. Columns 1 and 5 show that for both math and reading, first-year teachers are most responsive to changes in peer quality, as evidenced by the other experience group interactions having negative coefficients. The null hypothesis of equality of effects is rejected at traditional levels for math (p value = 0.001) but not for reading (p value = 0.54). For neither math nor reading are the marginal effects monotonic in experience. Teachers with between 10 and 24 years of experience are less responsive than teachers with more than 25 years of experience; however, one can reject the null hypothesis that teacher with more than 10 years of experience have the same marginal response as those with fewer than 10 years of experience at the 10 percent level for math (but not for reading). In sum, although the marginal effects are not monotonically decreasing in experience, for both math and reading the point estimates show

clear evidence that first-year teachers are more responsive to peer quality changes than teachers with 1 or more years of experience. Given that the first year of teaching is the one when teachers acquire the most on-the-job knowledge (as evidenced by the very steep experience value-added profile), these findings support a learning interpretation.

We also argue that teachers with greater labor-force attachment should be more responsive to peer quality changes. To test this prediction, we consider the interactions of regular licensure and of being certified with peer quality. Because teachers who obtain National Board certification and teachers who are fully certified have already invested the time and effort to obtain these qualifications, it is reasonable to assume that these are career teachers who intend to remain in teaching for several more years. If greater labor-force attachment increases the impetus to invest in on-the-job training, one might expect these career teachers to be more likely to invest in job-specific human capital (i.e., peer learning) and to be more responsive to peer quality changes than other teachers, on average. These interactions yield a consistent pattern and are generally supportive of this hypothesis. The interaction between peer quality and a fully licensed teacher is positive for both math and reading (in columns 2 and 6 of Table 8, respectively). The math result is statistically significant at the 1 percent level, while the reading effect is significant at the 10 percent level. Columns 3 and 7 present interactions with being fully certified. Although neither of these is statistically significant, the interactions between peer quality and being certified are positive for both math and reading (in columns 3 and 7, respectively). In sum, all the point estimates are consistent with the notion that teachers with greater labor market attachment are more sensitive to peer quality changes. Because only regular licensure results are statically significant, however, this prediction is only weakly supported by the evidence.

Because a direct learning story would involve acquiring knowledge from one's more knowledgeable peers, it is reasonable to assume that people may learn more from their peers when their peers are better than they are. Although teachers may also learn from their peers who know less than they do, it is reasonable to assume that more learning is likely to take place when a teacher is surrounded with peers who are knowledgeable than she is. One implication of this is that the marginal effect of peer quality should be smaller if one is the best among one's peers and higher when one is the worst. We consider whether there is a residual effect of being the best or

worst teacher in the grade after controlling for the average quality of one's peers.[17] All else equal, if a teacher learns from her peers, and if she is most likely to learn from her peers when they are better than she is, she should perform better when she is the worst of her peers and worse when she is the best of her peers (and has less to learn from them). We determine whether someone is the best or worst teacher in the grade based on whether she has the highest or lowest pre-period fixed effect. The results are presented in columns 4 and 8 of Table 8. We find that being the best teacher in the school and grade has a negative coefficient of -0.0067 for math and -0.0066 for reading, neither of which are statistically significant at traditional levels. Being the worst teacher in the school and grade has a positive but not statistically significant coefficient of 0.0002 for math and a positive and statistically significant 0.0136 for reading. The point estimates for both subjects are supportive of our hypothesis; however, the formal statistical tests that the effect of being the worst is the same being the best has a p value of 0.27 for math and 0.03 for reading. Because the marginal effect of mean peer quality is positive for both math and reading, the estimates indicate that peer effects are non-linear. An increase in peer quality for a teacher who goes from the best to the worst teacher yields a larger increase in own students' performance than the same increase in peer quality where the teacher was always the best among their peers. Because teachers arguably have more to learn from their peers who are better than them, this dynamic is exactly what one would expect if there were peer learning.

One of the principal differences between the alternative explanations for the peer effects observed in our main results is that if these peer effects are caused by learning, they should (1) be persistent over time and (2) have the same sign. Although differences in resources or motivation caused by having better peers should have little effect once the teacher's peers change, any learning that has occurred should stay with the teacher. We test whether peer quality continues to affect a teacher in future years. We include the first and second lag of the teacher's average peer quality, along with the contemporaneous measure of the quality of peers in that year. Results are presented in columns 3 and 4 of Table 9. Consistent with a learning interpretation, the peer effects are persistent. For math, in column 3, the coefficient on the contemporaneous effect is 0.031, the coefficient on the first lag is 0.037, and the coefficient on the second lag of the peer effect is 0.01. The contemporaneous effect and the lagged effect are both statistically

[17] One can also test this prediction by interacting the own value-added with the value-added of one's peers. This results in noise and statistically insignificant results.

significant at the 1 percent level, while the second lag is statistically significant at the 10 percent level. For reading, in column 4, the coefficient on the contemporaneous effect is 0.033, the coefficient on the first lag is 0.021, and the coefficient on the second lag of the peer effect is 0.018. All effects are statistically significant at the 5 percent level. For both subjects, the null hypothesis that the historical effects are equal to zero is rejected at the 5 percent level. These findings support a learning interpretation. It is important to note that including these lags suggests that the total effect of peers is larger than what we estimate in the main baseline regression. A one-standard-deviation increase in teacher peer quality that persists over time is associated with a 0.078 standard deviation increase in student test scores in math and a 0.072 standard deviation increase in student test scores in reading after two years.

Because learning is a cumulative process, another prediction of the learning model is that historical peer quality should "explain away" some of the predictive power of teacher fixed effects. If teachers learn from their peers (or as a result of exposure to their peers), and if learning becomes part of teacher ability, then there should be less variation attributable to the time-invariant teacher indicator variables conditional on the history of their peers. We test this hypothesis by comparing the fraction of the variance in test scores explained by the individual teacher effects in models that do and do not control for lagged peer quality. In models that include only contemporaneous peer quality, $Cov(\Delta A_{ij}, \theta_j) / Var(\Delta A_{ij})$ is 0.141 for math and 0.067 for reading. In models that include the first and second lags of peer quality, $Cov(\Delta A_{ij}, \theta_j) / Var(\Delta A_{ij})$ is 0.117 for math and 0.052 for reading. These differences suggest that between 18 and 24 percent of the contemporaneous own-teacher effect can be attributed to her peers in the two previous years. This suggests learning.[18] The persistence of peer quality over time is also important because it provides compelling evidence that our results are *not* driven by direct contact between students and their teacher's current peers. It also suggests that the observed spillovers are not due to transient changes in motivation or the allocation of non-classroom tasks

[18] As test of whether teachers acquire grade-specific knowledge such as how to teach fourth-grade math or more general teaching skills that would apply to all grades, we interacted the lagged peer value-added with indicators for whether the teacher moved to a new grade at the same school. If teachers acquired grade-specific skills, one would expect there to be greater persistence of peer effect for teachers who remain in the same grade. For math, the interactions are all positive and statistically insignificant, indicating that grade-specific knowledge may drive the spillovers for math. The joint hypothesis that all the interactions are equal to zero is rejected at the 10 percent level. For reading, however, the results are mixed. The second lag is less persistent, while the first lag is more persistent for mobile teachers. The joint hypothesis that all the interactions are equal to zero is rejected at the 5 percent level. In sum, the results of this test are mixed and inconclusive.

as a result of contemporaneous peer quality changes. Lastly, the persistence of the peer quality effects over time is further evidence that our central findings are not driven by dynamic student sorting.

Falsification Test: If it is learning, future peers should not matter.

As a final specification test, we augment the regressions in columns 3 and 4 of Table 9 to include a lead of peer quality from year t+1 in columns 5 and 6. Because teachers cannot learn from their future peers, future peers should have no effect on current student outcomes. This test of validity is also used in Mas and Moretti (2009). For both reading and math scores, the coefficient on mean peer quality the following year is not statistically significant, and the point estimates are smaller than those of either of the lags or the contemporaneous effects. The effect of lagged peers continues to be large and statistically significant for math, and the second lag remains large and statistically significant for reading. These results provide evidence that the effect of lagged peers is not a spurious correlation and suggests that the identification strategy is valid. The formal statistical test of the hypothesis that the future teacher effect is the same as the current teacher has a *p* value of 0.4 for math and 0.1 for reading. The test that the future teacher effect is the same as the lagged teacher has a *p* value of 0.00003 for math and 0.29 for reading. Lastly, the test that the future teacher effect is the same as the second lagged teacher has a *p* value of 0.87 for math and 0.11 for reading. This tests support the validity of our strategy for three reasons. First, the future teacher effects are not even marginally statistically significant. Second, the future teacher effects are smaller than the current effects and the lag effects for both math and reading, which under the null hypothesis of no causal effect and independence would happen only with probability $((0.5)^3)^2 = 0.0156$. Third, one can reject the null hypothesis that the future teacher effect is the same as at least one of the lags or contemporaneous effects for both subjects at the 10 percent level.

To summarize, all of the empirical predictions that suggest peer learning are supported by the data. Although not all of the interaction effects yield statistically significant estimates, all the point estimated are consistent with a peer learning interpretation of the spillovers. Although we cannot prove that the spillovers are due to peer learning, the evidence, taken in its entirely, suggests that teachers either learn directly from their peers (direct peer learning) or make the decision to invest in the learning as a direct result of exposure to better peers (peer induced learning).

Alternative Measures of Peer Quality

We use the mean of peer quality in most of our results because it is simple and likely to pick up some evidence of most types of peer effects. In this section, we try alternative approaches to measuring peer quality. We use the fixed effect associated with the best peer in the school and grade, the worst peer in the school and grade, the peer with the most similar level of experience, the peer with the most experience, and a peer of the same race. For comparison purposes, we also look at the effect of including the fixed effects associated with a random peer from the same school and grade. (*The random peer is taken from the teacher's actual peer group so it should have predictive power*.) Because many of the school grades in the dataset have only a few teachers, these measures may be identical for certain teachers. To the degree that they are the same in some instances, the exercise merely illustrates that the precise definition of peer quality chosen has little impact on our findings. We will, nevertheless, discuss some of the possible implications for the differences we do find.

The results for our alternative measures of peer quality are found in Table 10. None of the coefficients is as large as the effects for our preferred model (columns 3 and 6 of Table 5), suggesting that the mean of the teachers' fixed effects more accurately captures the relevant peer quality. On the other hand, all of the effects are positive, and some are statistically significant. Most of the effects, however, are similar to those found from using the fixed effect of a peer arbitrarily chosen from within the teacher's peer group in columns 6 and 12. The notable exception is the maximum and minimum of the fixed effects of the other peers in the same grade, school, and year in columns 1, 2, 7, and 8. For both math and reading, the effect of the maximum is larger than the effect of the minimum. For math, both of these effects are larger than for an arbitrary peer. The coefficients on the maximum are 0.0222 for math and 0.0164 for reading; both are significant at the 1 percent level. The coefficients on the minimum are 0.0197 for math and 0.0096 for reading. Because teachers are likely to learn from the best teacher rather than the worst teacher, the larger marginal effects of the quality of the best teacher than of the worst is consistent with the learning hypothesis.

None of the remaining columns shows any discernable patterns. Columns 3 and 9 show that teachers are no more responsive to the peer with the closest number of years of experience than a peer chosen at random from among the group of peers. Also, columns 4 and 10 show that teacher are no more responsive to peers of the same race than they are to other peers. These

results are not surprising give that the very small numbers of teachers in the average elementary school grade provide little opportunity to form cliques along racial or experience lines. With peer mentoring, the quality of the most experienced peer might be more important than that of a peer chosen at random from the group of peers. The results in columns 5 and 11 show no evidence of this. In sum, the mean quality of a teacher's peers appears to be the best measure of peer quality.

IX Conclusions

We document that a teacher's own performance is affected by the quality of her peers. In particular, *changes* in the quality of a teacher's colleagues (all other teachers in the same school who teach students in the same grade) are associated with *changes* in her students' test score gains. Using two separate measures of peer quality, one based on observable teacher qualifications and the other on estimated peer effectiveness, we find that teachers perform better when the quality of their peers improves within the same school over time. This within-teacher relationship is robust to including school-by-year fixed effects to account for changes in school attributes over time that could be correlated with changes in the make-up of the teacher population. Findings are also robust to including student fixed-effects. In our preferred model, we find that a one–standard-deviation improvement in *observable* teacher peer quality is associated with a 0.01 standard deviation increase in math and reading scores. Using estimated value-added (estimated out of sample to avoid simultaneity bias), which is a much better predictor of subsequent student achievement, we find that a one-standard-deviation improvement in *estimated* teacher peer quality is associated with a 0.0398 standard deviation increase in math scores and a 0.026 standard deviation increase in reading scores. This finding is robust to controlling for *observable* peer qualifications. We present a variety of falsification tests showing that our results are not driven by non-random dynamic student sorting across classrooms, or by the endogenous movement of teachers across grades within schools.

In an attempt to determine the mechanisms behind these spillovers, we test a variety predictions implied by different models. First, we show that less experienced teachers are generally more responsive to changes in peer quality than more experienced teachers, a finding that is consistent with the predictions of the model. We also find that teachers who are certified and have regular licensure are generally more responsive to peer quality, a finding that is consistent with the notion that career teachers have more of an incentive to invest in peer related

learning. We also find that the worst teachers in each school and grade perform better when their own value-added is the worst of all their peers than predicted and the best teachers in the school and grade perform worse when their own value-added is the best of all her peers. Insofar as teachers tend to learn more from peers who are better than themselves, this is also consistent with a peer learning story. The most compelling piece of evidence supporting the learning hypothesis is that the effect of teacher peer quality is persistent over time. Most peer effects that operate either through the education production function or through peer monitoring/pressure will have a contemporaneous effect. We show that for both math and reading, the quality of a teacher's peers the year before and even two years before affect her current students' achievement. For both subjects, the importance of a teacher's previous peers is as great as, or greater than, that of her current peers. The cumulative effect over three years of having peers with one standard deviation higher effectiveness is 0.078 standard deviations in math and 0.072 standard deviations in reading. Because teachers have about three peers on average, this is about one third of the size of the own-teacher effect, suggesting that over time, teacher peer quality is very important. Lastly, we find that peer quality in the previous two years "explains away" about one fifth of the explanatory power of individual teachers. This suggests that a sizable part of the own-teacher effect is learned as a result of exposure to her previous peers. Although we acknowledge that we cannot prove peer related learning, we believe these pieces of evidence lend themselves most naturally to a peer related learning interpretation (either learning directly from peers, or peer induced learning).

Although documenting productivity spillovers has been difficult, finding evidence of learning associated with peers in the workplace has been even more so. We document a systematic relationship between a teacher's peers' quality and her own performance (as measured by the performance of her students), and we present several pieces of evidence that are highly suggestive of peer learning. As a theoretical mater, knowledge spillovers are tremendously important in canonical models of economic growth, despite relatively little empirical support. Our findings provide important micro evidence of this type of productivity spillover. From a policy perspective, the finding that teachers learn as a result of their peers is important because it has direct implications for how teachers should be placed in schools and how they should be compensated. For example, compensation schemes that reward a teacher's performance relative to her peers may be detrimental to fostering peer learning. Also, the fact

that weaker and less experienced teachers are more responsive to peer quality than stronger and more experienced teachers suggests that novice teachers should be exposed to effective experienced teachers. This would imply that the high concentration of novice teachers in inner-city schools could be particularly detrimental to student performance at these schools in both the long and the short run.

Although we find little evidence, in our data, that a failure to account for contemporaneous peers lead to biased estimates of the effect of own-teachers on student test scores, we do show that the assumption of no spillovers across teachers is not valid. The fact that we document such spillovers should give pause to those who advocate using value-added estimates in policy and in teacher assessment and to researchers who make the assumption of a lack of spillovers.

Although our results are particularly relevant for the education setting, they add to the broader literature on peer effects. They highlight the type of data necessary to find evidence of peer effects and some of the features that may distinguish peer related learning from other types of peer spillovers. Although teachers in elementary school may be a somewhat unique group, the existence of peer effects and the suggestion of peer learning in this environment is suggestive that such spillovers may exist in other settings.

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics*, 25: 95-135.

Abowd, John., Robert Creecy, and Francis Kramarz. 2002. "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data." U.S. Census BureauTechnical Paper No. TP-2002-06.

Anderson, William, and Martin Wells. 2007. "Numerical Analysis in Least Squares Regression with an Application to the Abortion-Crime Debate." Paper presented at the 2nd Annual Conference on Empirical Legal Studies, City, ST.

Anderson, T.W., and Cheng Hsiao. 1981, "Estimation of dynamic models with error components", *Journal of the American Statistical Association*, vol.76, 598-606.

Angrist, Joshua, and Kevin Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." *American Economic Review*, 94(5): 1613-34.

Anthony, Emily, and Dan Goldhaber. 2007. "Can Teacher Quality Be Effectively Assessed? National Board Certification as a Signal of Effective Teaching." *The Review of Economics and Statistics*, 89(1): 134-50.

Azoulay, Pierre, Jialan Wang, and Joshua Graff Zivin. 2007. "Superstar Extinction." Unpublished.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2007. "Social Incentives in the

Workplace." Unpublished.

Battu, Harminder, Clive R. Belfield, and Peter J. Sloane. 2003. "Human Capital Spillovers within the Workplace: Evidence for Great Britain." *Oxford Bulletin of Economics and Statistics*, 65(5): 575-94.

Becker, Gary S. 1962. "Investment in Human Capital: A Theoretical Analysis." *Journal of Political Economy* 70: 9-49.

Bifulco, Robert, and Helen Ladd. 2005a Forthcoming. "The Impact of Charter Schools on Student Achievement: Evidence from North Carolina." *Education Finance Policy*.

Bifulco, Robert, and Helen Ladd. 2005b. "Race and Charter Schools: Evidence from North Carolina." Unpublished.

Brewer, Dominic, and Ronald Ehrenberg. 1994. "Do School and Teacher Characteristics Matter? Evidence from High School and Beyond." *Economics of Education Review*, 13(1): 1-17.

Brewer, Dominic, and Daniel Goldhaber. 2000. "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement." *Educational Evaluation and Policy Analysis*, 22(2): 129-45.

Clotfelter, Charles, Elizabeth Glennie, Helen Ladd, and Jacob Vigdor. 2008. "Teacher Bonuses and Teacher Retention in Low Performing Schools: Evidence from the North Carolina $1,800 Teacher Bonus Program." *Public Finance Quarterly*, 36(1): 63-87.

Clotfelter, Charles, Helen Ladd, and Jacob Vigdor. 2006. "Who Teaches Whom? Race and the Distribution of Novice Teachers." *Economics of Education Review*, 24(4): 377-92.

Clotfelter, Charles, Helen Ladd, and Jacob Vigdor. 2007. "How and Why Do Teacher Credentials Matter for Student Achievement?" National Bureau of Economic Research Working Paper No. 12828.

Clotfelter, Charles, Helen Ladd, Jacob Vigdor, and Justin Wheeler. 2007. "High Poverty Schools and the Distribution of Teachers and Principals." *North Carolina Law Review*, 85(5): 1345-80.

Cornelissen, Thomas. 2006. "Using Stata for a Memory Saving Fixed Effects Estimation of the Three-Way Error Component Model." FDZ Methodenreport 03/2006, IAB Nuernberg.

Costa, Dora, and Mathew Kahn. 2007. "Surviving Andersonville: The Benefits of Social Networks in POW Camps." *American Economic Review*, 97(4): 1467-87.

Duflo, Esther, and Emmanuel Saez. 2003. "The Role of Information and Social Interactions in Retirement Plan Decisions." *Quarterly Journal of Economics*, 118: 815-42.

Guryan, Jonathan, Kory Kroft, and Matt Notowidigdo. 2007. "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments." National Bureau of Economic Research Working Paper No. 13422.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2005. "Teachers, Schools and Academic Achievement." *Econometrica*, 73(2): 417-58.

Hastings Justine S., Richard Van Weelden, and Jeffrey Weinstein. 2007. "Preferences, Information, and Parental Choice Behavior in Public School Choice." National Bureau of Economic Research Working Paper No. 12995.

Hastings, Justine, and Jeffrey Weinstein. 2007. "No Child Left Behind: Estimating the Impact on Choices and Student Outcomes." National Bureau of Economic Research Working Paper No. 13009.

Hoxby, Caroline M. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." National Bureau of Economic Research Working Paper No. 7867.

Hoxby, Caroline M., and Gretchen Weingarth. 2005. "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects." Unpublished.

Ichino, Andrea, and Giovanni Maggi. 2000. "Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm." *Quarterly Journal of Economics*, 115(3): 1057-90.

Irwin, Douglas A., and Peter J. Klenow. 1994. "Learning-by-Doing Spillovers in the Semiconductor Industry." *Journal of Political Economy*, 102(6): 1200-27.

Jackson, C. Kirabo. 2008. "Student Demographics, Teacher Sorting and Teacher Quality: Evidence from the End of School Desegregation." Unpublished.

Jacob, Brian, and Lars Lefgren. 2008. "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education." *Journal of Labor Economics,25(1) : 101-136.*

Jovanovic, Boyan. 1979. "Job Matching and the Theory of Turnover." *Journal of Political Economy* 87: 972-90.

Kandel, Edward, and Edward Lazear. 1992. "Peer Pressure and Partnerships." *Journal of Political Economy*, 100(4): 801-17.

Kane, Thomas J., and Douglas O. Staiger. 2008. "Are Teacher-Level Value-Added Estimates Biased? An Experimental Validation of Non-Experimental Estimates." Unpublished.

Kane, Thomas, Douglas O. Staiger, and Stephanie Riegg. 2005. "School Quality, Neighborhoods and Housing Prices: The Impacts of school Desegregation." National Bureau of Economic Research Working Paper No. 11347.

Koedel, Cory. 2008. "An Empirical Analysis of Teacher Spillover Effects in Secondary School." http://ssrn.com/abstract=1260999.

Laschever, Ron. 2005. "The Doughboys Network: Social Interactions and Labor Market Outcomes of World War I Veterans." Unpublished.

Lavy, Victor, and Analia Schlosser. 2007. "Mechanisms and Impacts of Gender Peer Effects at School." National Bureau of Economic Research Working Paper No. 13292.

Lucas, Robert. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 22(1): 3-42.

McQuarrie, Frank, and Fred Wood. 1991. "Supervision, Staff Development, and Evaluation Connections." *Theory into Practice*, 30(2): 91-96.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies*, 60(3): 531-42.

Mas, Alexandre, and Enrico Moretti. 2009. "Peers at Work." forthcoming, *American Economic Review*

Moretti, Enrico. 2004a. "Human Capital Externalities in Cities." In *Handbook of Regional and Urban Economics*, ed. J. V. Henderson and J. F. Thisse, 1(4):, 2243-2291. -Elsevier.

Moretti, Enrico. 2004b. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review*, 94(3): 656-90.

Rockoff, Jonah. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review*, 94(2): 247-52.

Romer, Paul. 1990. "Endogenous Technological Change." *Journal of Political Economy*, 98(5, pt. 2): 71-102.

Rothstein, Jesse. 2007. "Do Value-Added Models Add Value? Tracking, Fixed Effects, and Causal Inference" Princeton University Working Paper..

Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results for Dartmouth

Roommates." *Quarterly Journal of Economics*, 116(2): 681-704.

Sorensen, Alan T. 2006. "Social Learning and Health Plan Choices." *RAND Journal of Economics*, 37(4): 929-45.

Thornton, Rebecca Achee, and Peter Thompson. 2001. "Learning from Experience and Learning from Others: An Exploration of Learning and Spillovers in Wartime Shipbuilding." *American Economic Review*, 91(5): 1350-68.

Todd, Petra, and Kenneth Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *The Economic Journal*, 113: F3-F33.

Wood, F., and J. Killion. 1998. "Job-Embedded Learning Matters to School Improvement." *Journal of Staff Development*, 19(1): 52-54.

Zimmerman, David J. 2003. "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment." *Review of Economics and Statistics*, vol. 85, issue 1, pages 9-23.

Zimmerman, Martin B. 1982. "Learning Effects and the Commercialization of New Energy Technologies: The Case of Nuclear Power." *Bell Journal of Economics*, 13(2), 297-310.

Table 1

Variable	Observations	Mean	Standard Deviation
Unit of Observation: Student-Year			
Math Scores	1361473	0.033	0.984
Reading Scores	1355313	0.022	0.984
Change in Math Score	1258483	0.006	0.583
Change in Reading Score	1250179	0.001	0.613
Black	1372098	0.295	0.456
White	1372098	0.493	0.500
Female	1372098	0.621	0.485
Parent Ed.: No HS Degree	1372098	0.107	0.309
Parent Ed.: HS Degree	1372098	0.428	0.495
Parent Ed.: Some College	1372098	0.315	0.464
Parent Ed.: College Degree	1372098	0.143	0.350
Same Race	1372098	0.649	0.477
Same Sex	1372098	0.496	0.500
Class Size	1372098	23.054	4.053
Unit of Observation: Teacher-Year			
Experience	91243	12.798	9.949
Experience 0	92511	0.063	0.242
Experience 1 to 3	92511	0.165	0.371
Experience 4 to 9	92511	0.230	0.421
Experience 10 to 24	92511	0.365	0.481
Experience 25+	92511	0.164	0.371
Teacher Exam Score	92511	-0.012	0.812
Advanced Degree	92511	0.197	0.398
Regular Licensure	92511	0.670	0.470
Certified	92511	0.039	0.194
Peer Experience 0	85490	0.064	0.164
Peer Experience 1 to 3	85490	0.166	0.255
Peer Experience 4 to 9	85490	0.230	0.289
Peer Experience 10 to 24	85490	0.364	0.334
Peer Experience 25+	85490	0.164	0.256
Peer Teacher Exam Score	85490	-0.009	0.578
Peer Advanced Degree	85490	0.198	0.274
Peer Regular Licensure	85490	0.676	0.426
Peer Certification	85490	0.039	0.140

Note: The few teachers with more than 50 years of experience are coded as having 50 years of experience.

Table 2

	Panel A: Selected Summary Statistics by Teacher's Status in Previous Year				Panel B: Difference Between Selected Characteristics of Movers and Peers	
	Same Grade and School	Same School, Different Grade	Different School	New to Data	Same School, Different Grade	Different School
Percentage of All Teachers	65.82	5.95	7.38	20.85		
Experience	14.62	13.05	11.82	6.70	-0.008	-0.272
	(9.68)	(9.16)	(9.18)	(9.28)	(0.208)	(0.186)
Teacher Exam Score	-0.02	0.02	-0.03	0.05	0.018	-0.012
	(0.82)	(0.79)	(0.78)	(0.70)	(0.017)	(0.015)
Advanced Degree	0.23	0.22	0.21	0.15	0.007	0.016
	(0.42)	(0.41)	(0.41)	(0.36)	(0.009)	(0.008)*
Regular Licensure	0.52	0.50	0.54	0.36	0.029	0.05
	(0.50)	(0.50)	(0.50)	(0.48)	(0.005)**	(0.004)**
Certified	0.07	0.07	0.06	0.02	0.005	0.003
	(0.25)	(0.25)	(0.23)	(0.14)	(0.006)	(0.005)

Panel A: Standard deviations in parentheses.
Panel B: Standard errors in parentheses. +, *, and ** indicate significance of a *t* test that the mean is equal to zero at 10%, 5%, and 1%, respectively.

Table 3

Effect of Teacher Peer Quality on Math Achievement

	Dependent Variable: Mathematics Test Score			Dependent Variable: Reading Test Score		
	1	2	3	4	5	6
Model	School Fixed Effects	Student-School Fixed Effects	Teacher-School and School-Year Fixed Effects	School Fixed Effects	Student-School Fixed Effects	Teacher-School and School-Year Fixed Effects
	OLS	OLS	OLS	OLS	OLS	OLS
Lagged Score	0.7674	—	0.7658	0.739	—	0.7332
	[0.0021]**	—	[0.0018]**	[0.0016]**	—	[0.0016]**
Experience 1 to 3	0.0651	0.1005	0.0547	0.0408	0.0616	0.0324
	[0.0045]**	[0.0082]**	[0.0041]**	[0.0035]**	[0.0070]**	[0.0038]**
Experience 4 to 9	0.0816	0.1215	0.0683	0.0547	0.0743	0.0323
	[0.0046]**	[0.0080]**	[0.0054]**	[0.0035]**	[0.0069]**	[0.0050]**
Experience 10 to 24	0.0997	0.1383	0.0747	0.0754	0.0967	0.0377
	[0.0045]**	[0.0078]**	[0.0070]**	[0.0035]**	[0.0066]**	[0.0064]**
Experience 25+	0.1025	0.1368	0.0616	0.0835	0.1008	0.0295
	[0.0048]**	[0.0084]**	[0.0088]**	[0.0037]**	[0.0071]**	[0.0080]**
Peer Experience 1 to 3	0.0248	0.042	0.0288	0.0071	0.0204	0.017
	[0.0071]**	[0.0142]**	[0.0069]**	[0.0055]	[0.0117]+	[0.0064]**
Peer Experience 4 to 9	0.0193	0.0363	0.0264	0.006	0.0153	0.0132
	[0.0073]**	[0.0145]*	[0.0074]**	[0.0056]	[0.0120]	[0.0068]+

Peer Experience 10 to 24	0.0247	0.0442	0.0218	0.0161	0.0303	0.0294
	[0.0072]**	[0.0142]**	[0.0075]**	[0.0055]**	[0.0117]**	[0.0069]**
Peer Experience 25+	0.0238	0.0383	0.0209	0.0145	0.0259	0.0154
	[0.0078]**	[0.0152]*	[0.0083]*	[0.0059]*	[0.0125]*	[0.0075]*
Licensure Score	0.0172	0.0179	—	0.0043	0.0018	—
	[0.0012]**	[0.0031]**	—	[0.0009]**	[0.0022]	—
Advanced Degree	-0.0057	-0.0018	—	-0.004	-0.0015	—
	[0.0024]*	[0.0073]	—	[0.0019]*	[0.0050]	—
Regular Licensure	0.0375	0.0583	—	0.0215	0.0324	—
	[0.0041]**	[0.0084]**	—	[0.0032]**	[0.0068]**	—
Certified	0.0347	0.0477	—	0.0156	0.0207	—
	[0.0046]**	[0.0111]**	—	[0.0035]**	[0.0081]*	—
Peer Licensure Score	0.0007	0.0027	0.0034	-0.0008	0.0013	-0.0022
	[0.0020]	[0.0037]	[0.0024]	[0.0015]	[0.0030]	[0.0022]
Peer Advanced Degree Share	0.0031	0.0016	0.0049	-0.0038	-0.0025	-0.0099
	[0.0038]	[0.0074]	[0.0043]	[0.0029]	[0.0061]	[0.0040]*
Peer Regular Licensure Share	0.0092	0.0112	-0.0096	0.0113	0.0128	-0.0024
	[0.0064]	[0.0124]	[0.0066]	[0.0050]*	[0.0103]	[0.0060]
Peer Certification Share	0.0126	0.0355	0.0025	0.0017	0.0191	-0.008
	[0.0069]+	[0.0133]**	[0.0076]	[0.0054]	[0.0111]+	[0.0068]
Observations	1200633	1200633	1200633	1192896	1192896	1192896
R-squared	0.16	0.5		0.16	0.49	

Robust standard errors clustered by school-teacher in brackets.

* Significant at 10%; * significant at 5%; ** significant at 1%.

Note: All models include indicator variables for the gender and racial matches between the teacher and the students, class size, and grade by year fixed effects. All regressions include student demographic control variables except models that include student fixed effects. All regressions include teacher control variables except models that include teacher fixed effects (teacher experience included in all models). The omitted teacher experience group is teachers with zero years of experience. All regressions include an indicator variable for having missing experience data, and control for the proportion of peers with missing experience data.

Table 4

	School Fixed Effects	Student Fixed Effects	Teacher-School and School-Year Fixed Effects	School Fixed Effects	Student-School Fixed Effects	Teacher-School and School-Year Fixed Effects
	1	2	3	4	5	6
	OLS	OLS	OLS	OLS	OLS	OLS
Bad Peers	-0.025	-0.044	-0.025	-0.01	-0.022	-0.009
	[0.007]***	[0.013]***	[0.007]***	[0.005]*	[0.011]*	[0.006]
Good Peers	0.011	0.018	-0.004	0.012	0.022	0.01
	[0.005]**	[0.009]**	[0.005]	[0.003]***	[0.007]***	[0.005]**
Cov(Peers chars 1, Y)/Var(Y)	0.0065	0.0105	0.0061	0.0062	0.0084	0.0062
Cov(Peers chars 2, Y)/Var(Y)	0.0042	0.0065	0.0045	0.0047	0.0054	0.0028
Cov(Peers chars 3, Y)/Var(Y)	0.0040	0.0069	0.0038	0.0035	0.0055	0.0055

Prob(Teacher peers don't matter)	0.00	0.02	0.07	0.00	0.05	0.00
Prob(Teacher peers don't matter other than exp.)	0.21	0.27	0.27	0.16	0.50	0.30
Prob(Teacher peers' experience doesn't matter)	0.01	0.04	0.04	0.01	0.09	0.00

Robust standard errors clustered by school-teacher in brackets.
* Significant at 10%; * significant at 5%; ** significant at 1%.
Note: Results based on regression coefficients in Table 3. Bad peers shows the difference between having peer characteristics one standard deviation below the mean and having peer characteristics at the mean. Good peers shows the difference between having peer characteristics one standard deviation above the mean and having peer characteristics at the mean. Bad peers have experience 0 compared to experience of 4-9 years. Good peers have experience 10-24 years compared to 4-9 years. Peers chars 1 is the value added of all peer characteristics, Peers chars 2 is the value-added of teacher peers experience variables and Peers chars 3 is the value-added of the non-experience teacher peer variables..

Table 5

Effect of Mean Peer Value-Added on Student Achievement: 2001 Through 2006 Sample

Dependent Variable:	Math Test Score			Reading Test Score		
	1	2	3	4	5	6
	School Fixed Effects	Student Fixed Effects	Teacher-School and School-Year Effects	School Fixed Effects	Student Fixed Effects	Teacher-School and School-Year Effects
	OLS	OLS	OLS	OLS	OLS	OLS
Lagged Score	0.7728	—	0.7712	0.7293	—	0.7233
	[0.0009]**	—	[0.0009]**	[0.0010]**	—	[0.0010]**
Teacher Effect	0.1268	0.1689	—	0.0547	0.0785	—
	[0.0031]**	[0.0062]**	—	[0.0027]**	[0.006]**	—
Mean Teacher Peers Effect	0.0522	0.0604	0.0398	0.0262	0.0346	0.026
	[0.0037]**	[0.0076]**	[0.0049]**	[0.0035]**	[0.0044]*	[0.0050]**
Observations	684696	689387	684696	679262	683850	679262
R-squared	0.18	0.88		0.17	0.87	

Robust standard errors clustered by school-teacher in brackets.
+ Significant at 10%; * significant at 5%; ** significant at 1%.

Note: Estimated using data from 2001 through 2006. All models include indicator variables for the gender and racial matches between the teacher and the students, class size, and year-by-grade fixed effects. All regressions include student demographic control variables except models that include student fixed effects. All regressions include teacher control variables except models that include teacher fixed effects (note that teacher experience is included in all models). All models include indicators for missing estimated value-added as well as the proportion of peers with no estimated value-added. The omitted teacher experience group is teachers with missing experience data.

Table 6

Test for Systematic Sorting Bias

	1 Math	2 Reading
Teacher Effect (same subject)	0.12	0.052
	[0.006]**	[0.005]**
Future Teacher Effect (same subject)	0.002	-0.002
	[0.003]	[0.004]
Grade Effects	YES	YES
School by Year Effects	YES	YES
Pr(P>|t|) Future effect =Current effect	<0.001	<0.001
Pr(P>|t|) Future teacher effect=0	0.38	0.64
Observations	231390	229507
R-squared	0.22	0.19

Robust standard errors in brackets.
+ Significant at 10%; * significant at 5%; ** significant at 1%.

Table 7

Predictors of Receiving a New Peer

	1 New Teacher from Same School, Different Grade	2 New Teacher, Different School	3 New Teacher Not from State Data	4 New Teacher from Same School, Different Grade	5 New Teacher, Different School	6 New Teacher Not from State Data	7 New Teacher from Same School, Different Grade	8 New Teacher, Different School	9 New Teacher Not from State Data
Lag Mean Math Test Score Growth	-0.001	-0.01	-0.01	-0.001	-0.018	-0.013	-0.013	-0.012	-0.002
	[0.011]	[0.010]	[0.014]	[0.013]	[0.011]	[0.016]	[0.016]	[0.015]	[0.018]
Lag Mean Reading Test Score Growth	0.01	0.006	0.003	0.009	0.012	0.009	0.02	0.01	-0.001
	[0.010]	[0.010]	[0.013]	[0.012]	[0.011]	[0.016]	[0.014]	[0.014]	[0.017]
Lag %teachers: 0 years experience	-0.003	0.029	-0.173	-0.003	0.03	-0.171	0.006	0.116	-0.165
	[0.095]	[0.105]	[0.132]	[0.095]	[0.105]	[0.132]	[0.107]	[0.116]	[0.147]
Lag %teachers: 1 to 3 years experience	0.051	-0.01	-0.23	0.051	-0.009	-0.228	0.072	0.074	-0.21
	[0.086]	[0.097]	[0.120]+	[0.086]	[0.097]	[0.120]+	[0.095]	[0.103]	[0.128]
Lag %teachers: 4 to 9 years experience	0.052	-0.021	-0.201	0.052	-0.02	-0.199	0.066	0.065	-0.203
	[0.086]	[0.095]	[0.119]+	[0.086]	[0.095]	[0.119]+	[0.094]	[0.100]	[0.127]
Lag %teachers: 10 to 24 years experience	0.053	-0.046	-0.262	0.053	-0.046	-0.26	0.05	0.023	-0.269
	[0.085]	[0.095]	[0.121]*	[0.085]	[0.095]	[0.120]*	[0.094]	[0.101]	[0.128]*
Lag %teachers: 24+ years experience	0.105	-0.028	-0.245	0.105	-0.027	-0.243	0.118	0.045	-0.235

	[0.088]	[0.098]	[0.121]*	[0.088]	[0.098]	[0.121]*	[0.096]	[0.105]	[0.128]+
Lag Average Math Test Score	—	—	—	0.001	-0.012	-0.012	-0.003	-0.01	-0.004
	—	—	—	[0.012]	[0.011]	[0.014]	[0.014]	[0.013]	[0.016]
Lag Average Reading Test Score	—	—	—	-0.001	0.014	0.005	0.004	0.012	0.002
	—	—	—	[0.012]	[0.011]	[0.014]	[0.014]	[0.014]	[0.016]
Lag Mean Teacher Value-Added Math	—	—	—	—	—	—	-0.013	-0.016	-0.041
	—	—	—	—	—	—	[0.022]	[0.021]	[0.027]
Lag Mean Teacher Value-Added Reading	—	—	—	—	—	—	0.003	-0.027	-0.001
	—	—	—	—	—	—	[0.033]	[0.025]	[0.035]
Observations	19550	19550	19550	19550	19550	19550	12466	12466	12466
R-squared	0.6	0.63	0.68	0.6	0.63	0.68	0.6	0.62	0.68

Robust standard errors in brackets.
+ Significant at 10%; * significant at 5%; ** significant at 1%.
Includes grade fixed effects and school by year fixed effects.

Table 8

Interaction of Peer Quality and Own Characteristics

	Dependent Variable: Math Test Score				**Dependent Variable: Reading Test Score**			
	1	2	3	4	5	6	7	8
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Peer Effect	0.0599	0.0205	0.0376	0.0354	0.063	0.0198	0.0376	0.0163
	[0.0283]*	[0.0069]**	[0.0059]**	[0.0067]**	[0.0249]*	[0.0078]*	[0.0059]**	[0.0063]**
Peer Effect*Experience 1 to 3	-0.003	—	—	—	-0.0412	—	—	—
	[0.0273]	—	—	—	[0.0254]	—	—	—
Peer Effect*Experience 4 to 9	-0.0077	—	—	—	-0.039	—	—	—
	[0.0294]	—	—	—	[0.0264]	—	—	—
Peer Effect*Experience 10 to 24	-0.0361	—	—	—	-0.0384	—	—	—
	[0.0293]	—	—	—	[0.0257]	—	—	—
Peer Effect*Experience 25+	-0.016	—	—	—	-0.0278	—	—	—
	[0.0310]	—	—	—	[0.0272]	—	—	—
Peer Effect*Regular Licensure	—	0.0323	—	—	—	0.0167	—	—
	—	[0.0074]**	—	—	—	[0.0086]+	—	—
Peer Effect*Certified	—	—	0.0129	—	—	—	0.0189	—
	—	—	[0.0143]	—	—	—	[0.0183]	—
Best Teacher	—	—	—	-0.0067	—	—	—	-0.0066
	—	—	—	[0.0049]	—	—	—	[0.0046]
Worst Teacher	—	—	—	0.0002	—	—	—	0.0136
	—	—	—	[0.0049]	—	—	—	[0.0046]**
Prob(Worst = Best)	—	—	—	.27	—	—	—	.03-**??**
Prob(TFX(exp >10) = (TFX(exp >10))	0.08	—	—	—	.21	—	—	—
Teacher-School Effects	YES	YES	YES	YES	YES	YES	YES	YES
School-Year Effects	YES	YES	YES	YES	YES	YES	YES	YES
Observations	684752	684752	684752	684752	679230	679230	679230	679230

Robust standard errors clustered by school-teacher in brackets.
+ Significant at 10%; * significant at 5%; ** significant at 1%.
Note: Estimated using data from 2001 to 2006. The variable "peer effect" is the mean estimated value-added of a teacher's peers (all other teachers at the same school in the same grade during the same year). All models include indicator variables for the gender and racial matches between the teacher and the students, class size, student demographic control variables, teacher experience, indicators for missing estimated value-added, the proportion of peers with no estimated value-added, and year-by-grade fixed effects. The omitted teacher experience group is teachers with zero years of experience.

Table 9

Effect of Historical Peers on Current Student Test Scores

	1 Math	2 Reading	3 Math	4 Reading	5 Math	6 Reading
Peer Effect (own subject)	0.038	0.028	0.031	0.033	0.019	0.035
	[0.005]**	[0.005]**	[0.0081]**	[0.0086]*	[0.0102]*	[0.0123]**
Lagged Peer Effect (own subject)	-	-	0.037	0.021	0.051	0.025
	-	-	[0.0074]**	[0.0084]*	[0.0093]**	[0.0112]*
Second Lag of Peer Effect (own subject)	-	-	0.01	0.018	0.011	0.0308
	-	-	[0.0066]+	[0.0073]*	[0.0090]	[0.0089]**
Lead of Peer Effect (own subject)	-	-	-	-	0.009	0.007
	-	-	-	-	[0.0094]	[0.0103]
Teacher-School Effects	YES	YES	YES	YES	YES	YES
School-Year Effects	YES	YES	YES	YES	YES	YES
Observations	680479	678389	374478	371640	252538	250652
Cov(Y,TFX)/Var(Y)	0.141	0.067	0.117	0.052	0.117	0.052
Prob(Future = current)	-	-	-	-	0.4	0.1
Prob(Future = lag)	-	-	-	-	<000	0.29
Prob(Future = second lag)	-	-	-	-	0.87	0.11

Robust standard errors clustered by school-teacher in brackets.
+ Significant at 10%; * significant at 5%; ** significant at 1%.
Note: Estimated using data from 2001 to 2006. The variable "peer effect" is the mean estimated value-added of a teacher's peers (all other teachers at the same school in the same grade during the same year). All models include indicator variables for the gender and racial matches between the teacher and the students, class size, student demographic control variables, teacher experience, indicators for missing estimated value-added, the proportion of peers with no estimated value-added, and year-by-grade fixed effects. The omitted teacher experience group is teachers with missing experience data.

Table 10

Different Peer Specifications												
Dependent Variable	Math Test Score Growth						Reading Test Score Growth					
	1	2	3	4	5	6	7	8	9	10	11	12
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Minimum Peer	0.0197	-	-	-	-	-	0.0096	-	-	-	-	-
	[0.0035]**	-	-	-	-	-	[0.0039]*	-	-	-	-	-
Maximum Peer	-	0.0222	-	-	-	-	-	0.0164	-	-	-	-
	-	[0.0038]**	-	-	-	-	-	[0.0041]**	-	-	-	-
Closest Experience Peer	-	-	0.0123	-	-	-	-	-	0.0055	-	-	-
	-	-	[0.0031]**	-	-	-	-	-	[0.0034]+	-	-	-
Mean Peer	-	-	-	0.0422	-	-	-	-	-	0.031	-	-
	-	-	-	[0.0124]**	-	-	-	-	-	[0.0144]*	-	-
Mean Own Race	-	-	-	0.0164	-	-	-	-	-	-0.0119	-	-
	-	-	-	[0.0117]	-	-	-	-	-	[0.0134]	-	-
Most Experienced Peer	-	-	-	-	0.0178	-	-	-	-	-	0.0141	-
	-	-	-	-	[0.0033]**	-	-	-	-	-	[0.0038]**	-
Random Peer	-	-	-	-	-	0.0143	-	-	-	-	-	0.007
	-	-	-	-	-	[0.0025]**	-	-	-	-	-	[0.0033]*
Teacher-School Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
School-Year Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	684752	684752	684752	620360	684752	684752	679230	684752	679230	615406	679230	679230

Robust standard errors clustered by school-teacher in brackets.
+ Significant at 10%; * significant at 5%; ** significant at 1%.

Note: The dependent variables denote the peer for which the estimated fixed effect is used. For example, minimum peer is the teacher with the lowest estimated value-added among a teacher's peers, while random peer is the estimated effect of a peer chosen at random (based on a random number generator). Estimated using data from 2001 to 2006. The variable "peer effect" is the mean estimated value-added of a teacher's peers (all other teachers at the same school in the same grade during the same year). All models include indicator variables for the gender and racial matches between the teacher and the students, class size, student demographic control variables, teacher experience, indicators for missing estimated value-added, the proportion of peers with no estimated value-added, and year-by-grade fixed effects. The omitted teacher experience group is teachers with missing experience data.

Appendix

Appendix Table 1

Regression Estimates from First Stage Teacher Value-Added Estimation

	-1 Math	-2 Reading		1 Cont'd Math	2 Cont'd Reading
Grade 4	-0.895	-0.734	Class Size	-0.005	-0.002
	[0.017]**	[0.018]**		[0.000]**	[0.000]**
Grade 5	-0.872	-0.722	Teacher: 1-3 years experience	0.068	0.03
	[0.018]**	[0.019]**		[0.018]**	[0.019]
Student Male	0	0.009	Teacher: 4-10 years experience	0.078	0.04
	[0.003]	[0.003]**		[0.018]**	[0.018]*
Student Black	-0.072	-0.082	Teacher: 10-24 years experience	0.071	0.033
	[0.007]**	[0.008]**		[0.018]**	[0.019]+
Student Hispanic	-0.023	0.004	Teacher: 25+ years experience	0.057	0.019
	[0.009]**	[0.009]		[0.019]*	[0.020]
Student American Indian	-0.099	-0.069	School: %Black	0.063	0.229
	[0.011]**	[0.012]**		[0.061]	[0.064]**
Student Mixed Ethnicity	-0.086	-0.059	School: %White	0.105	0.28
	[0.011]**	[0.012]**		[0.060]+	[0.063]**
Student White	-0.105	-0.074	School: %Hispanic	0.416	0.47
	[0.007]**	[0.008]**		[0.085]**	[0.089]**
Parental Education: Some High School	-0.002	0.001	School: %Free-Lunch Eligible	0.067	0.025
	[0.003]	[0.003]		[0.016]**	[0.017]
Parental Education: High School Graduate	0.006	0.007	School: Log Enrollment	0.001	0.002
	[0.004]	[0.004]+		[0.007]	[0.008]
Parental Education: Some College	0.002	0.008	School: Urban Fringe (Large City)	-0.068	-0.062
	[0.003]	[0.004]*		[0.018]**	[0.019]**
Parental Education: Prof. Graduate School	0.01	0.006	School: Mid-Sized City	-0.064	-0.07
	[0.003]**	[0.003]		[0.017]**	[0.018]**
Parental Education: Junior College Graduate	0.026	0.007	School: Urban Fringe (Mid-Sized City)	-0.077	-0.06
	[0.004]**	[0.004]+		[0.018]**	[0.019]**
Parental Education: College	0.037	-0.014	School: Large Town	-0.032	-0.033
	[0.008]**	[0.008]+		[0.030]	[0.031]
Parental Education: Graduate School	-0.016	0.312	School: Small Town	-0.075	-0.072
	[0.565]	[0.593]		[0.019]**	[0.020]**
Teacher and Student Are Same Race	0.009	0.006	School: Rural (Inside CBSA)	-0.088	-0.075
	[0.003]**	[0.003]*		[0.018]**	[0.019]**
Teacher and Student Are Same Sex	0.006	-0.004	School: Rural (Outside CBSA)	-0.042	-0.049
	[0.003]*	[0.003]		[0.018]*	[0.019]**
			Observations	535332	533060

Standard errors in brackets. + significant at 10%; * significant at 5%; ** significant at 1%.

Note: These models are estimated using student data from the years 1995 through 2000. All regressions include year fixed effects. The reference student's ethnic group is Asian students. The reference parental education group is no high school. The reference city size category is large city. The omitted teacher experience group is teachers with zero years of experience.

Appendix Note 1: Estimating Teacher Fixed Effects

There are several specifications used in the literature to estimate teacher value-added; however, the predictive power of estimated teacher fixed effects are generally robust to the chosen specification [Kane and Staiger (2008)]. We estimate teacher fixed effects using the adjusted test score growth model described in Section II. Specifically, teacher effectiveness comes from estimation of equation [3] using data from 1995 through 2000.

$$A_{it} = \hat{\delta} A_{it-1} + \phi X_{it} + \varsigma Z_{st} + \eta W_{jt} + \theta_j + \xi_{gt} + \varepsilon_{ijgst} \qquad [3]$$

All variables are defined as before, with the addition of θ_j ,which is the effect of teacher *j*. $\hat{\delta}$ is the coefficient on lagged test scores in a test score growth model obtained from a 2SLS regression using the second lag of test scores as an instrument for lagged test scores. Because we use the first year of data to compute test score growth for 1996, the actual estimation sample used spans the years 1996 through 2000. Because we need estimates of teacher value-added that can be comparable across schools, grades, and classes, we *do not* include school or student fixed effects but rather include a set of demographic controls for the students and schools.[19]

Researchers have pointed out that there is substantial measurement error in test scores such that the coefficient on lagged test score would be downward biased.[20] Under the assumption that measurement errors in test scores are not correlated over time, , many researchers have used the second lag of test scores as an instrument for the lagged test score [as proposed in Anderson and Hsiao (1981) and Todd and Wolpin (2003)]. One downside of this approach is that it requires several years of data for each student and is impractical to implement in models that include large vectors of three fixed factors. We propose a method that builds on this solution but allows one to use more of the available data. The basic idea is that if using the second lag of test scores as an instrument for the lagged test score results in a consistent estimate of δ , then one can use this estimate to adjust the test score growth outcome variable for the full sample and obtain consistent estimates on the coefficient for other characteristics that may be correlated with lagged test scores. We present a proof of this below.

This is implemented by first estimating the instrumental variables regressions on the full sample, where the second lag of test scores is used as an instrument for the first lag of test scores. The consistent estimates of the coefficient on lagged test scores is therefore estimated *in sample*. Because this can be estimated only for students with two lags of test scores, this 2SLS model uses only grade 5 outcomes. We estimate the sample analog of equation [4] (replacing δ with $\hat{\delta}$) using all the observations for which lagged test score are available. As a practical

[19] Specifications that include student or school fixed effects identify teacher value-added based on within-school or within-student variation. If teachers are very different across schools, then much of the variation in teacher quality (i.e., the cross-school variation) will be absorbed by the school fixed effect, making estimated effects across schools impossible to compare. Including student fixed effects further exacerbates this problem by allowing only comparisons of teachers who teach the same groups of students. If those teachers who teach the gifted and talented students are of different average quality from those who teach the regular students, the estimated teacher value-added can be used only to compare teachers who share the same students, so that comparing teachers who teach different students (even within the same school) may be misguided.

[20] This is the same as saying that there is attenuation bias on the coefficient of lagged test scores in [3] due to measurement error in test scores. If lagged test scores are correlated with other covariates (very likely), this will bias the coefficients for all covariates.

matter, although the 2SLS coefficient on lagged test scores (between 0.97 and 0.95) is *much* smaller than the OLS estimates (between 0.70 and 0.76), the peer effects results are similar across models, so that our results are not driven by any modeling assumptions from this procedure. However, the teacher value-added estimates perform better in the falsification test of section VII (as would be expected if the 2SLS adjustment removes measurement error bias from the teacher estimates).

Proof : Consider the following. We can rewrite [3] as $A_{it} - \delta A_{it-1} = \phi H_{it} + \varepsilon_{it}$ where H_{it} denotes all observable covariates and teacher, grade, and school subscripts are suppressed. Suppose we have a consistent unbiased estimate $\hat{\delta}$ of δ such that $\lim_{n\to\infty} \hat{\delta} = \delta$ and $E(\hat{\delta}) = \delta$. Where test scores are measured with error such that $\hat{A}_{it} = A_{it} + u_{it}$ and one uses $\hat{\delta}$ in the place of δ, this can be written as [4] below.

$$A_{it} - \hat{\delta} A_{it-1} = \phi H_{it} - (\delta - \hat{\delta}) A_{it-1} + \delta u_{it-1} - u_{it} + \varepsilon_{it} \qquad [4]$$

Equation [4] can be directly estimated using OLS where the unobserved error term is $\delta u_{it-1} - u_{it} + \varepsilon_{it}$. Because H_{it} is uncorrelated with u_{it-1}, u_{it}, and ε_{it} by assumption, the OLS estimate $\hat{\phi}$ of ϕ from [4] will be unbiased and consistent *iff* $E[\widehat{Cov}(H_{it}, (\delta - \hat{\delta}) A_{it-1})] = 0$ where $\widehat{Cov}$ is the sample covariance. Using Slutsky's theorems, because $\lim_{n\to\infty} \hat{\delta} = \delta$ and $E(\hat{\delta}) = \delta$, it follows that $\lim_{n\to\infty}[\widehat{Cov}(H_{it}, (\delta - \hat{\delta}) A_{it-1})] = 0$ and $E[\widehat{Cov}(H_{it}, (\delta - \hat{\delta}) A_{it-1})] = 0$ so that $\lim_{n\to\infty}(\hat{\phi}) = \phi$ and $E(\hat{\phi}) = \phi$ from [4].