

How Do Teachers Improve? The Relative Importance of Specific and General Human Capital

Ben Ost *

Cornell University
Department of Economics
bw292@cornell.edu

November 29, 2009

Abstract

One of the most consistent findings in the literature on teacher quality is that teachers improve with experience, especially in the first several years. This study extends this research by separately identifying the effect of general teaching experience and grade-specific experience. Using within-teacher variation, I find that both general experience and grade-specific experience improve teacher performance. For math scores, the magnitude implies that teachers who always repeat grade assignments improve approximately 35% faster than teachers who never repeat grade assignments. In addition to furthering our understanding of how teachers improve with experience, this paper contributes to a literature on task-specific human capital.

*I would like to thank Ron Ehrenberg, Kirabo Jackson, Jordan Matsudaira, Doug Webber, Ken Whelan, James Cowan and Carrie Ost for very helpful comments on earlier drafts. Also, I am indebted to the North Carolina Education Research Data Center for making available the confidential data used in this study.

1 Introduction

Like many professions, successful teaching requires a variety of skills and knowledge. For example, the skill set of a good fourth-grade elementary school teacher includes pedagogical knowledge, fourth grade curriculum familiarity, classroom management skills, and communication skills. The economic literature on teacher human capital acquisition has found compelling evidence that teachers improve with experience; however, which skills develop with experience and which skills are most important for effective teaching have not been investigated. This paper contributes to our understanding of teacher improvement by separately identifying the importance of acquiring course-specific human capital and general teaching human capital. Course-specific human capital encompasses areas like lesson planning and curriculum familiarity, whereas general teaching human capital includes skills applicable to teaching any course.

To address this question, I use administrative matched teacher-student data that follows each teacher in the North Carolina public school system from 1995-2007. Results from a value-added model show that teachers improve with experience and the magnitude of these improvements depends upon the frequency with which they are able to apply both general and grade-specific human capital.¹ In my preferred specification which includes teacher-by-school fixed effects, grade-specific experience is found to be approximately one third as important as general experience for benefiting student math scores. There is little evidence that grade-specific experience benefits students' reading scores. Controlling for the level of grade-specific experience, general teaching experience benefits both math and reading scores.

While this paper is the first to examine specific and general experience in the context of teaching, several studies have attempted to separately identify task-specific experience from general experience. The major advantage of the present paper is the ability to control not only for firm and year fixed effects, but also individual worker fixed effects. In addition, unlike the firm/industry ef-

¹Value-added models rely on there being no systematic student sorting according to the teacher characteristic of interest. While previous research such as Rothstein (Forthcoming) and Clotfelter et al. (2006) show evidence of students sorting into classrooms, I find little evidence that this sorting is systematically related to experience within a teacher. I explore these issues in detail in the falsification section.

fects literature, the changes in task assignment that I exploit are extremely common and I provide evidence that phenomena such as Ashenfelter dips do not occur prior to a grade re-assignment. Furthermore, because teachers frequently change grade assignments, I am able to control for both the act of switching itself as well as the number of switches each teacher has experienced. My preferred estimates are thus identified off of the specific pattern and order in which grades are taught.

This study's contributions are threefold. First, it provides direct empirical evidence that within an occupation, task-specific human capital acquisition can significantly affect productivity. Second, it provides some evidence as to which skills teacher professional development programs should focus on. Lastly, the results of this paper can guide policy regarding teacher grade assignments.

2 Literature

2.1 Task-Specific Human Capital

While there is a wealth of empirical literature devoted to examining the relative importance of firm-specific and general human capital, relatively little research has empirically analyzed the role of task-specific human capital in wage formation. In a theoretical paper on task-specific human capital, Gibbons and Waldman (2004) explain that “some of the human capital an individual acquires on the job is specific to the tasks being performed, as opposed to being specific to the firm” (p. 203). They argue that task-specific human capital has broad applicability compared to firm- or industry-specific human capital and task-specific human capital can theoretically explain phenomena such as cohort effects, job design, and promotion patterns. Unfortunately, relatively few empirical studies have examined the role of task-specific human capital, possibly because most longitudinal data lacks descriptions of job tasks.

In a recent paper, Gathmann and Schönberg (2007) examine worker transition using a panel dataset of German workers combined with a cross-sectional survey of job tasks. Matching each

occupation to one of three “task requirements” they find that task-specific tenure can account for between 25-40% of wage growth. Since their paper assumes that within an occupation tasks are uniform, the analysis is restricted to observing individuals who change jobs. The essence of the identification is comparing individuals who move to a job with a new set of task requirements to individuals who move to a job with similar task requirements as their previous employment. The authors note, however, that a potential drawback to this approach is the possibility of unobserved differences between workers who move to a similar occupation and workers who move to a dissimilar occupation. A major advantage of Gathmann and Schönberg (2007) over my paper, however, is that the authors examine a representative sample of occupations whereas this paper is limited to teachers.

Another study analyzes the role of task-specific familiarity in the context of financial analyst forecasting performance (Clement et al., 2007). Unlike Gathmann and Schönberg, this paper uses cross sectional analysis and tests whether analysts with more experience analyzing firm restructurings are more accurate in analyzing future restructurings than are other analysts. This paper has the advantage of examining a narrowly defined type of specific experience and using a direct measure of productivity; however, because the data used is cross-sectional, unobserved heterogeneity across analysts cannot be ruled out. Importantly, the results found in this study may be driven by differential attrition among analysts.²

My paper is similar to Clement et al. (2007) in that my analysis is restricted to a single occupation and a narrowly defined “specific experience.” Also, I observe productivity directly rather than assuming wages are equal to marginal product. Given the strong theoretical justification for a divergence between wages and marginal products over the life course (Lazear, 1979) observing productivity is a more direct test of the role of specific human capital. Unlike Clement et al. (2007), however, my data is longitudinal and thus I am able to control for unobserved time-invariant heterogeneity.

²If analysts that perform poorly on their first firm-restructuring assignment are fired, then currently employed analysts with previous experience are likely better on average than currently employed analysts with no firm-restructuring experience.

2.2 Teacher Experience

While the impact of many teacher characteristics is still debated, there exists an emerging consensus that teacher experience positively contributes to student learning. In a meta-analysis of the value-added literature, Harris (2008) finds that eight of nine studies that use value-added models show evidence of teachers improving with experience.

However, the only paper of which I am aware that explores a mechanism for *how* teachers improve is Jackson and Bruegmann (2009). The authors show that teachers improve when exposed to higher quality peers, thus demonstrating that part of teacher improvement is based on learning from other teachers. My paper builds on this research by identifying the type of skills that are most important to learn.

3 Data

I use longitudinal administrative data that links students to their teachers in the state of North Carolina between 1995-2007.³ This data includes detailed information on student, classroom, teacher, and school characteristics as well as a standardized measure of math and reading achievement for students in grades three through eight. For each student, the data include race, gender, parental education, free or reduced lunch status, and test scores for each grade. Available teacher characteristics include gender, race, highest degree earned, years of teaching experience, undergraduate institution, and licensure test scores.⁴ Years of teaching experience is based on the number of years credited to a teacher for the purposes of salary calculation and thus should reflect all experience in any district.

By matching teacher information to classroom records, I am able to identify the grade taught by each teacher in each year. Using this information I construct a variable indicating the number of years a teacher has previously taught her current grade assignment. Because middle and high

³This data has been extensively cleaned and standardized by the North Carolina Education Research Data Center housed at Duke University.

⁴This data is described in great detail in Clotfelter et al. (2007).

school teachers often teach multiple grades simultaneously, and because student test score data is most complete for third through fifth grade, I restrict my sample to elementary teachers who teach self-contained single-grade classes.

While the North Carolina data includes a link between student test scores and teachers, the teacher listed is actually the proctor of the student exam, not necessarily the classroom teacher. For elementary classrooms, the proctor is likely to be the classroom teacher, but to improve the accuracy of teacher-student matches, I limit the sample to confidently matched students. Following Clotfelter et al. (2007), Rothstein (Forthcoming), and Jackson and Bruegmann (2009) I consider a proctor to be the classroom teacher so long as the teacher's grade assignment matches the grade of the proctored exam. In addition, I drop cases where a proctor administered more than half of his/her tests to a different grade level.⁵

To ensure comparability across years, I standardize all test scores by grade and year.⁶ In order to implement some econometric specifications, I require a lagged test score in addition to current test scores. Students who are only present in the data for a single year are therefore dropped. The exception is for third graders, since the lagged third grade test is actually given to students at the beginning of the third grade rather than in second grade. While the data includes complete teacher data starting from 1995, complete student data is only available starting in 1997. I use the 1995-1996 period to calculate grade-specific experience, but only use 1997-2007 in regressions.

As discussed by Koedel and Betts (2008), achievement tests that contain ceilings may lead to systematic measurement error since students near the ceiling are unable to make further gains. I test for a ceiling in the North Carolina data by comparing a kernel density of each distribution to that of the normal density. Figures 1 and 2 demonstrate that there is little reason to be concerned with ceiling effects for these tests.

⁵In describing this data, Jackson and Bruegmann (2009) note that "According to state regulation, the tests must be administered by a teacher, principal, or guidance counselor. Discussions with education officials in North Carolina indicate that tests are always administered by the students' own teachers when these teachers are present. 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." Given these procedures, it is unlikely that a teacher assigned to teach a given grade would proctor an exam for other students in the same grade.

⁶Standardization is made prior to any data restrictions and thus the mean and standard deviation of the analysis sample is not exactly zero and one respectively.

3.1 Data Limitation: Grade-Specific Experience

The data include information on teaching experience accrued before the sample period; however, my measure of grade-specific experience is limited to the sample time frame. For example, a teacher with ten years experience in 2003 accrued the latter eight years during the sample frame, but the data provides no information regarding the grades she taught in her first two years (1993-1994). Thus, I cannot exactly determine this teacher's grade-specific experience for any year. In general, I cannot exactly calculate grade-specific experience for teachers who have pre-sample experience.

I address this data limitation with two distinct approaches. The first approach simply restricts the data to teachers who began teaching during the sample period and whose grade assignments are thus fully observed. The second approach uses all teachers for analysis and imputes grade-specific experience accrued prior to the sample period. This imputation is implemented by assuming that the histogram of grades a teacher taught out-of-sample matches the average histogram of grades taught by other teachers in his or her school. Each approach has benefits and drawbacks. Restricting the sample to fully observed teachers leads to an unrepresentatively inexperienced sample, while imputing grade-specific experience leads to measurement error in an explanatory variable. Because the measurement error introduced is not necessarily classical however, the measurement error may bias estimates and it is difficult to assess its direction and magnitude.

Given that imputation may lead to biased estimates, my preferred specification restricts the sample to fully observed teachers.⁷ The fact that this sample of teachers is unusually inexperienced is likely a minor issue since previous research has found that most improvement occurs during the first several years (Rivkin et al., 2005).

Table 1 shows descriptive statistics for both the full sample and the sample which is restricted to fully observed teachers. As can be seen from this table, the restricted sample has considerably less experience on average than the full sample.⁸ In addition, 25.6% of the full sample of teachers

⁷While my preferred specifications restrict the sample to fully observed teachers, results based on the full sample are very similar and are shown in the appendix.

⁸The substantial difference in average experience is not surprising. By definition, teachers in the restricted sam-

have an advanced degree whereas only 13.1% of the restricted sample of teachers have an advanced degree. The upper panel of Table 1 shows that restricting the sample to relatively inexperienced teachers also leads to a slightly different sample of students. Students in the restricted sample perform worse than students in the full sample and these students are also more likely to be a minority. These differences reflect that fact that schools with weaker, minority students, have relatively high teacher turnover rates and thus are disproportionately staffed by recently hired teachers. While the restricted sample of teachers is clearly not representative of teachers as a whole, it is the complete universe of recently hired teachers in the state of North Carolina and thus interesting in and of itself. The next section provides a more nuanced description of the relationship between grade-specific experience and general experience in this sample.

4 General Experience vs Grade-Specific Experience

In the absence of grade assignment changes, experience and grade-specific experience would be perfectly collinear and I would only be able to identify a single effect. To investigate the prevalence of grade assignment changes, Table 2 presents a transition matrix showing grade assignments in year $t+1$ as a function of grade assignment in year t . This table demonstrates that approximately 30 percent of teachers switch grade assignments after teaching third, fourth or fifth grade. This table also documents that teachers are much more likely to switch to adjacent grades than distant grades.⁹

The frequent switching documented in Table 2 leads to a substantial divergence between experience and grade-specific experience. Table 3 presents a cross tabulation of grade-specific experience and experience. Approximately a third of teachers teach a new grade in their second year of teaching, and less than a third teach the same grade five times in their first five years teaching. This pattern continues in later years and suggests that it is possible to separately identify grade-specific

ple must have less than 13 years of experience whereas the full sample includes many teachers with 30+ years of experience.

⁹Evaluating whether experience in adjacent grades is more beneficial than experience in distant grades is hindered by the fact that relatively few teachers acquire experience in distant grades.

and general experience for this sample.

4.1 Grade-Specific Experience and Student Performance

As a preliminary analysis of the effect of grade-specific experience, I perform simple mean comparisons of average student performance. Figure 3 graphically shows changes in average student test score gains as grade-specific experience varies. Each panel in this figure holds absolute years of experience constant and graphs the average student test score gains for teachers with various levels of grade-specific experience. These figures show that teachers with more grade-specific experience perform better in terms of their students' test score gains.

This relationship is especially clear for lower experience levels and is more pronounced for math score gains than reading score gains. While these figures are suggestive, they simply reflect raw correlations and by themselves cannot be interpreted as implying a causal relationship. However, in a later section I obtain a more controlled estimate of the impact of grade-specific experience which confirms the implications of the simple average comparisons.

5 Identification Strategy

When using teacher fixed effects, the effect of grade-specific experience is identified by two different sources of variation. First, grade-specific experience diverges from general experience whenever a teacher receives a new grade assignment. Second, conditional on the number of times a teacher switches grades, grade-specific experience varies depending on the pattern of grades taught and the order in which these grades are taught.

5.1 Test for Systematic Grade Switching

Since one source of variation is based on grade switching, estimates which use this source of variation may be biased if teacher switching is correlated with expected performance.¹⁰ If teachers are switched to grades in which they have less experience in years when one expects that they will do particularly poorly, this may lead to overstating the importance of grade-specific experience. To test whether teacher switching is related to expected performance, I estimate equations (1) and (2). Since it is impossible to directly measure “expected” performance, I test whether current performance predicts whether a teacher is switched the following year.¹¹

$$\mathbf{1}(g_{j,t} = g_{j,t+1}) = \lambda \sum_g \left| N_{gs(t+1)} - N_{gs(t)} \right| + \zeta \bar{A}_{j_t g_t s_t t} + \beta E_{jt} + \omega_{js} + \xi_g + \phi_t + \epsilon_{jt} \quad (1)$$

$$\mathbf{1}(g_{j,t} = g_{j,t+1}) = \lambda \sum_g \left| N_{gs(t+1)} - N_{gs(t)} \right| + \zeta \bar{\Delta A}_{j_t g_t s_t t} + \beta E_{jt} + \omega_{js} + \xi_g + \phi_t + \epsilon_{jt} \quad (2)$$

The variable $\mathbf{1}(g_{j,t} = g_{j,t+1})$ is an indicator that is unity when teacher j repeats grade assignments and zero when teacher j switches grade assignments. The variable $\bar{A}_{jgst}$ denotes the average test scores for students taught by teacher j in grade g in school s in time t .¹² The vector E_{jt} includes experience and grade-specific experience in period t . Both specifications include teacher-by-school, grade, and year fixed effects denoted by ω_{js} , ξ_g and ϕ_t respectively. The coefficient λ captures the impact of school-by-year changes in the demand for teachers of each grade.¹³

¹⁰More exactly, estimates may be biased if teacher switching is correlated with expected performance conditional on all observables.

¹¹Since principals might consider raw test scores in addition to test score progress, I define performance as absolute test score in equation (1) and as test score gains in equation (2). I use a linear probability model (LPM) to predict whether a teacher switches grade assignments between years. This model is run as a linear probability model rather than a non-linear model for simplicity and because empirically, predicted probabilities all lie between zero and one using the LPM.

¹²These regressions are run at the classroom level because the tests aim to capture principals’ responses to class performance. When the same regressions are estimated at the student level, the results are similar but the magnitudes of the estimates are considerably smaller.

¹³Hoxby (2000) exploits population variation to identify the effect of class size on student achievement. Similarly, population variation leads to changes in the number of classes per grade. When a particularly large cohort of students passes through a school, teachers may need to be switched around each year in order to create extra sections for the large cohort. The variable $N_{gs(t)}$ denotes the number of sections of grade g in school s in time t . Thus, $\sum_g \left| N_{gs(t+1)} - N_{gs(t)} \right|$ gives the total number of section changes in a school for a given year. I include this variable because as the number of section changes increase, I expect that teachers are more likely to switch grades to fill those

In performing this test, I am primarily interested in the coefficient ζ because this coefficient reflects the extent to which teachers are switched due to their current students' performance.¹⁴

As can be seen in the first two columns of Table 4, there is weak evidence that teachers are switched based on current performance. A one standard deviation improvement in average reading test gains leads to a 0.9 percentage point decrease in the probability of being switched. There is a small negative effect from math gains as well; however this is not statistically significant. The level of average reading scores and math scores are not statistically significant predictors of teacher switching. Although the evidence of systematic teacher switching is rather weak, I cannot conclude that estimates identified from teacher switches will be completely free of bias.

5.2 Test for Systematic Changes in Grade-Specific Experience

The second (and more exact) source of identifying variation is based upon the magnitude of grade-specific experience changes. For example, when a teacher switches grade assignments, this generally leads to a decrease in grade-specific experience; however, grade-specific experience sometimes increases as a result of a switch.¹⁵ When controlling for both whether a teacher switches and current grade-specific experience, estimates will be biased if next year's grade-specific experience is correlated with expected performance. I test for this correlation by examining whether current teacher performance predicts next year's grade-specific experience. It should be emphasized that this test cannot detect correlations between grade-specific experience and expected performance that are unrelated to current performance. I estimate:

$$EXPGRD_{j,t+1} = \gamma \bar{A}_{j_t g_t s_t t} + \beta E_{jt} + \omega_{js} + \xi_g + \phi_t + \epsilon_{jt} \quad (3)$$

$$EXPGRD_{j,t+1} = \gamma \overline{\Delta A}_{j_t g_t s_t t} + \beta E_{jt} + \omega_{js} + \xi_g + \phi_t + \epsilon_{jt} \quad (4)$$

empty positions.

¹⁴If one teacher switches grades, another teacher will possibly need to switch grades as well, thus standard errors are clustered at the school-year level. In practice, using OLS standard errors or other clustering levels lead to similar results.

¹⁵In my data, 13.45% of switches result in increases in grade-specific experience.

The variable $EXPGRD_{t+1}$ denotes the grade-specific experience for teacher j in time $t + 1$ and all other variables are defined as in (1) and (2). As shown in the third and fourth column of Table 4 there is little evidence that grade-specific experience changes are related to current performance. Because there is some evidence that teacher switches are correlated with performance, my empirical model controls both for switching last period and the total number of switches a teacher has experienced. Based on the above tests, I conclude that conditional on these switches, the resulting grade-specific experience variation is not systematically related to previous teacher performance.

6 Empirical Model

To evaluate the impact of teacher characteristics on student outcomes I use a value-added model (VAM) that controls for student characteristics, teacher characteristics, and several fixed effects to predict future test scores. My preferred specification controls for the lag of test score; however, I explore the robustness of results across other value-added models in the appendix.¹⁶

$$A_{ijgst} = \alpha A_{ij_{t-1}g_{t-1}s_{t-1},t-1} + \beta X_i + \delta C_{ijgst} + \rho V_{ij} + \pi D_{jt} + f(Exp_{jt}) + g(Expgrd_{jt}) + \xi_g + \omega_{js} + \phi_t + \epsilon_{ijgst} \quad (5)$$

A_{ijgst} is the test score of student i taught by teacher j in grade g in school s in time t . The student characteristic vector X_i includes student gender, ethnicity, subsidized lunch status, and parental education. Classroom characteristics such as class size and average peer characteristics (excluding student i) are denoted by C_{ijgst} . The vector V_{ij} includes interactions between the student and teacher ethnicity and sex.¹⁷ The vector D_{jt} includes a control for whether the teacher switched in the previous period and the total number of past switches a teacher has experienced. This model includes grade, teacher-by-school and year fixed effects denoted by ξ_g , ω_{js} and ϕ_t respectively.

¹⁶The lagged test score VAM is used in many recent studies including Aaronson et al. (2007), Kane et al. (2006), Jackson and Bruegmann (2009), and others. Controlling for the lag of test score is found to outperform other value-added methodologies in an experimental validation study by Kane and Staiger (2008). In their paper Kane and Staiger find no evidence that non-experimental value-added estimates are biased, relative to experimental estimates.

¹⁷Dee (2005) shows that gender and ethnicity match may effect student achievement.

Experience and grade-specific experience enter through $f(\cdot)$ and $g(\cdot)$. While some past studies have controlled for national board certification status, I do not control for this variable because national board certification itself may reflect teacher quality gained from experience.

As has been noted in previous research, measurement error in lagged test scores can bias estimates on all coefficients. I follow the procedure suggested by Anderson and Hsiao (1981) and Todd and Wolpin (2003) and use the second lagged test score as an instrument for the first lagged test score. Generally, this IV specification would drop any student who lacks three consecutive test scores leading to an unrepresentative sample that disproportionately represents students in relatively stable situations. I use the estimator proposed by Jackson and Bruegmann (2009) which avoids this significant data restriction. Essentially the Jackson and Bruegmann estimator uses the restricted student sample to estimate α using the double lag as an instrument for the lag of test score. The estimate of α is then used in estimating equation (5) for the entire sample.¹⁸

While Rothstein (Forthcoming) demonstrates significant non-random sorting of students into classrooms, this will only bias my results to the extent that this sorting is correlated with grade-specific teacher experience within a teacher. Furthermore, because I control for absolute years of teaching experience, estimates of the impact of grade-specific experience will only be biased if students are sorted into classrooms depending on the teachers' grade-specific experience conditional on a fixed level of overall teaching experience. I explore these concerns in the falsification section and find little evidence that within a teacher, students are differentially sorted as the teacher gains experience or grade-specific experience.

Following Rockoff (2004), Aaronson et al. (2007), Koedel and Betts (2007), and others, I model experience effects as a cubic polynomial. Given the nearly perfect collinearity between general experience and year effects when teacher fixed effects are included, it is necessary to assume that teacher experience has no impact on quality after a certain threshold of experience. As in Rockoff (2004), Harris and Sass (2007), and Koedel and Betts (2007) I use 10 years as the cutoff for general experience.¹⁹

¹⁸See the online appendix of Jackson and Bruegmann (2009) for a proof of the consistency of this estimator.

¹⁹See Rockoff (2004) for a discussion and justification of this assumption. In practice, this assumption is imple-

7 Results

Results from the above specification are shown in Table 5 and Table A-1. Consistent with previous research, a teacher's experience is found to positively impact student outcomes. While controlling for number of years of teaching (general experience), grade-specific experience also has a positive impact on student math scores.

Based on the restricted sample with teacher-by-school fixed effects, estimates for math scores imply that a teacher who teaches the same grade for each of her first five years helps students perform 0.137 standard deviations better than students with a novice teacher. If, instead, a teacher teaches different grades every year for her first five years, she helps students perform 0.0985 standard deviations better than a novice teacher. In other words, a teacher who always repeats grade-assignments improves 39% more than a teacher who never repeats grade assignments. Estimates based on the full sample are similar, and are shown in the appendix.

Based on the restricted sample there is no statistically significant effect of grade-specific experience on reading scores. Estimates based on the full sample, however, show small statistically significant benefits of grade-specific experience for reading scores. However, as is shown in Table (A-4), the results for reading scores are somewhat sensitive to the exact value-added model chosen, so I am hesitant to conclude that grade-specific experience benefits reading scores. While I have no definitive explanation as to why grade-specific experience matters for math but not reading, one possible cause is the fact that similar reading skills are taught in each grade whereas math curricula change dramatically for each grade.²⁰

To examine the entire experience profile, Figure 4 and Figure 5 plot predicted outcomes for teachers as they gain experience. These figures graphically present the functions shown in the first

mented by recoding the experience variable so that teachers with above ten years of experience are made to have exactly ten years of experience. In the unrestricted sample, this recoding leads to grade-specific experience exceeding general experience for the approximately 2% of teacher-year observations that have grade-specific experience above 10 years. By definition, experience must be greater or equal than grade-specific experience so I impose a cutoff of 10 years on grade-specific experience as well.

²⁰The North Carolina standard curriculum "five competency goals" demonstrate this point. Between third and fifth grades, all five reading competency goals remain identical for each grade whereas all five math competency goals change for each grade (North Carolina Department of Education, 2009).

and fourth columns of Table 5.²¹ For math scores, the experience profile is concave, increasing for both grade-specific experience and general experience. The effect of grade-specific experience ranges from 30-40% of the general experience effect depending on the experience level. As was previously discussed, reading scores improve with teacher general experience but not with grade-specific experience.

The second, third, fifth and sixth columns of Table 5 show that results are similar when using school fixed effects or teacher-by-school-by-grade fixed effects. The model with school fixed effects has the advantage of utilizing a larger amount of variation to identify the grade-specific experience effect, since teachers who never switch grades still contribute to these estimates. Unfortunately, this model may be biased if teachers have inherent traits that make them both frequently switched and inherently weak or strong teachers. Furthermore, differential attrition may lead to unobserved differences between experienced and inexperienced teachers.²²

Models with teacher-by-school-by-grade fixed effects have the advantage of controlling for unobserved differences in a teacher's inherent ability to teach different grades; however, these estimates are identified only for teachers who switch grades and then switch back to their original grade.²³ Since only 7.48% of teachers switch grades and then switch back, estimates that include teacher-by-school-by-grade fixed effects are identified from a relatively small fraction of teachers.

Given that results are qualitatively similar when including school, teacher-by-school, or teacher-by-school-by-grade fixed effects, it appears that the exact source of identifying variation is relatively unimportant. That being said, my preferred specification includes teacher-by-school fixed effects because it controls for possibly important unobserved heterogeneity without severely restricting the identifying sample.

²¹Standard error bands are calculated using the formula $Var(aE + bE^2 + cE^3) = E^2Var(a) + E^4Var(b) + E^6Var(c) + 2E^3Cov(a, b) + 2E^4Cov(a, c) + 2E^5Cov(b, c)$. This formula explains why bands "fan out" and become large for high levels of experience. The larger standard errors for higher levels of experience is fairly mechanical and is not due to small number of observations for higher levels of experience.

²²Hanushek et al. (2005), Boyd et al. (2008), and Goldhaber et al. (2007) all find evidence that less effective teachers are more likely to leave a school.

²³When a teacher switches away from a grade and later returns, grade-specific experience is not colinear with general experience within a teacher-grade. If a teacher only teaches grades in continuous runs, experience and grade-specific experience are always perfectly colinear within each teacher-grade.

8 Falsification

A major concern in using value-added models to measure teacher productivity is the extent to which students are non-randomly assigned to teachers. Importantly, the non-random assignment of students to teachers is necessary but not sufficient to bias estimates of experience effects. In order for non-random student assignment to bias experience effects, student assignment must be systematically related to teacher experience within a teacher. Just as random student assignment is sufficient for eliminating bias in the presence of non-random teacher placement, randomly assigning teachers (according to their experience) is sufficient for eliminating biased estimates of experience effects, in the presence of non-random student sorting.

Rothstein (Forthcoming) provides evidence that students are non-randomly assigned to teachers; however, since I include teacher-fixed effects, bias is only created if students are sorted differentially over a teacher's tenure. In addition, as demonstrated by Rothstein (2009), the lagged test score model that I use has the advantage that it can account for dynamic systematic sorting based on observable past student performance. Specifications that include student-fixed effects and exclude the lagged score are unable to control for dynamic student sorting.

With this type of bias in mind, I test for systematic assignment based on grade-specific experience. Following Jackson and Bruegmann (2009), I test for this type of bias by examining whether a future teacher's experience is correlated with current student outcomes, conditional on observables. While I confirm Rothstein's finding that a student's future teacher appears to have an "effect" on current performance, I find weak evidence that this non-random sorting is related to grade-specific experience.²⁴

Table 6 shows the results of a regression in which the experience levels of each student's future teacher is included in the regression given by equation (5). This regression examines whether a student's future teacher's experience levels are correlated with a student's current performance, conditional on observables. As can be seen in this table, there is limited evidence of non-random

²⁴This indicates that my estimates of grade-specific experience effects are likely not affected by non-random sorting, but individual teacher fixed effects should not be relied upon to evaluate that specific teacher's effectiveness.

sorting based on teacher grade-specific experience. One specification shows that teachers are given students with slightly better reading scores as they gain grade-specific experience, but this result is not robust across models. Given 36 coefficient estimates, finding one to be marginally significant could easily be due to random chance.

Because of the nature of the falsification exercise, students must be observed in three consecutive years to be included in the regression. This restriction leads to a considerable sample size reduction which lowers the power of these regressions. It is possible that estimates given in Table 6 would be statistically significant if the sample size was equivalent to Table 5. As shown in the appendix, when using the full sample of teachers (with grade-specific experience imputations), there is some evidence that students are sorted according to a teacher’s general experience level for math scores. Regardless, the magnitudes of these false experience “effects” are quite small compared to the estimates given in Table 5. Based on this rather weak evidence of student sorting, I conclude that it is unlikely that previous estimates are substantially biased due to student sorting.

9 Conclusion

The fact that teachers improve with experience is commonly cited as one reason why teacher attrition is problematic. This paper shows that reassigning a teacher to a new grade has consequences similar to teacher attrition because his or her grade-specific human capital is wasted. While it is very difficult and expensive to affect teacher attrition through policy, improving teacher grade assignments is more straightforward to implement. Based on conversations with principals and teachers, it is apparent that completely avoiding grade assignment switches is unrealistic. Teachers are sometimes switched due to changes in enrollment, personality clashes amongst teachers in a grade, or when a teacher is perceived to feel “too comfortable” with their current grade. In cases where grade reassignments are unavoidable, however, principals should consider providing teachers who are new to their grade assignment many of the supports provided to teachers who are

generally inexperienced.²⁵

More generally, the evidence presented in this paper suggests that the applicability of human capital is directly related to the similarity between past and future tasks performed. Thus, in analyzing the effect of worker movement, the most relevant question may not be whether the worker moves to the same industry or firm, but rather whether the new job requires similar tasks to the old. Gathmann and Schönberg (2007) provide some evidence of the importance of task-specific human capital in the context of job changes, but this literature remains under developed.

Given that teachers gain both general and grade-specific human capital, principals should make a concerted effort to increase the frequency with which teachers are able to repeat grade assignments, especially in the first few years of teaching. Approximately one third of teachers in my sample teach two different grades in their first two years and I see little theoretical or empirical justification for these assignments. Given the political and financial costs of most reforms aimed at improving educational outcomes, education policy makers should take advantage of the relatively easy and cheap efficiency gain of giving teachers repeat grade assignments.

²⁵ A small percentage of schools have consciously decided to frequently change grade assignments based on their belief that grade-specific human capital is less important than student-teacher relationships that develop over the course of a year. These schools have teachers “follow” their students and therefore teacher grade assignments change incrementally each year. While one could theoretically evaluate the effectiveness of such an approach, in practice, too few students in North Carolina are exposed to this system. In the student-teacher matched sample, fewer than 0.1 percent of students who were in the sample from third to fifth grade had the same teacher all three years.

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.
- Anderson, T. W., and Cheng Hsiao. 1981. Estimation of dynamic models with error components. *Journal of the American Statistical Association*.
- Boyd, Donald, Pam Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2008. Who leaves? teacher attrition and student achievement. NBER Working Papers 14022, National Bureau of Economic Research, Inc.
- Clement, Michael B., Lisa Koonce, and Thomas J. Lopez. 2007. The roles of task-specific forecasting experience and innate ability in understanding analyst forecasting performance. *Journal of Accounting and Economics*, 44(3), 378–398.
- Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2006. Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, 41(4), 778.
- Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2007. Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26(6), 673–682.
- Dee, Thomas S.. 2005. A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95(2), 158–165.
- Gathmann, Christina, and Uta Schönberg. 2007. How general is human capital? a task-based approach. IZA Discussion Papers 3067, Institute for the Study of Labor (IZA).
- Gibbons, Robert, and Michael Waldman. 2004. Task-specific human capital. *American Economic Review*, 94(2), 203–207.
- Goldhaber, Dan, Betheny Gross, and Daniel Player. 2007. Are public schools really losing their "best"? CRPE working paper 20072, Center on Reinventing Public Education.

- Hanushek, Eric A., John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin. 2005. The market for teacher quality. NBER Working Papers 11154, National Bureau of Economic Research, Inc.
- Harris, Douglas. 2008. The policy uses and policy validity of value-added and other teacher quality measures. In D. H. Gitomer (Ed.), *Measurement Issues and the Assessment of Teacher Quality*. SAGE Publications.
- Harris, Douglas N., and Tim R Sass. 2007. Teacher training, teacher quality and student achievement. CALDER Working Papers 3, CALDER.
- Hoxby, Caroline M.. 2000. The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics*, 115(4), 1239–1285.
- Jackson, Clement Kirabo, and Elias Bruegmann. 2009. Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, 1(4), 85–108.
- Kane, Thomas J., and Douglas O. Staiger. 2008. Are teacher-level value-added estimates biased? an experimental validation of non-experimental estimates. Nber working papers, National Bureau of Economic Research, Inc.
- Kane, Thomas J., Douglas O. Staiger, and Stephanie K. Riegg. 2006. School quality, neighborhoods, and housing prices. *American Law and Economics Review*.
- Koedel, Cory, and Julian Betts. 2007. Re-examining the role of teacher quality in the educational production function. Working Papers 0708, Department of Economics, University of Missouri.
- Koedel, Cory, and Julian Betts. 2008. Value-added to what? how a ceiling in the testing instrument influences value-added estimation. Working Papers 0807, Department of Economics, University of Missouri.
- Lazear, Edward P. 1979. Why is there mandatory retirement? *Journal of Political Economy*, 87(6), 1261–84.

- North Carolina Department of Education. 2009. Accountability services. <http://www.dpi.state.nc.us/accountability/testing/eog/>.
- Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458.
- Rockoff, Jonah E.. 2004. The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2), 247–252.
- Rothstein, Jesse. 2009. Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, 4(4), 537–571.
- Rothstein, Jesse. Forthcoming. Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, (14442).
- Todd, Petra E., and Kenneth I. Wolpin. 2003. On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, 113(485), F3–F33.

Tables, Figures and Appendix

Figure 1: Distribution of Math Scores

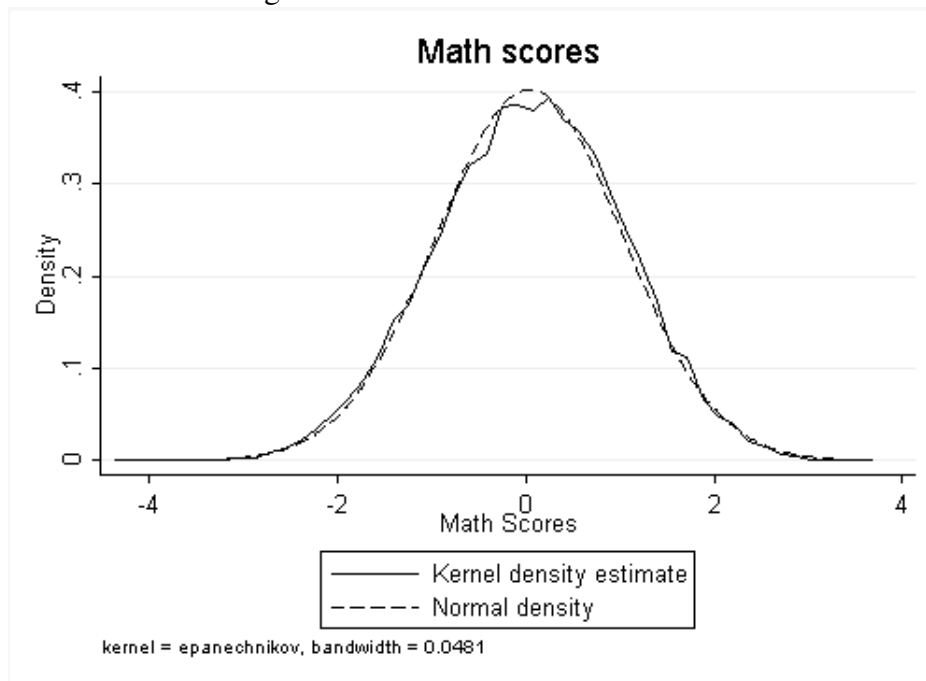
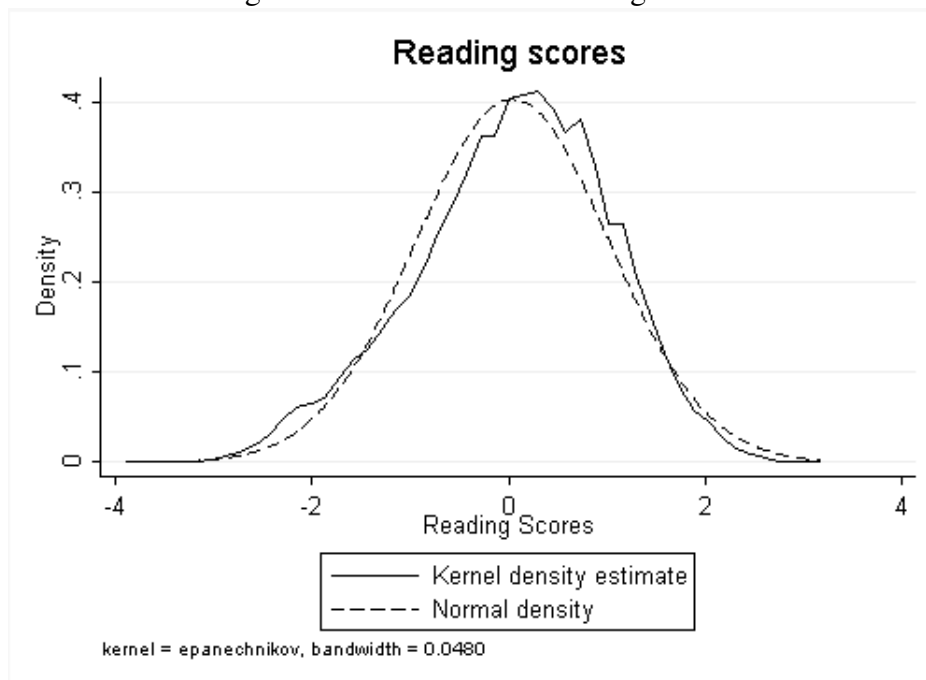


Figure 2: Distribution of Reading Scores



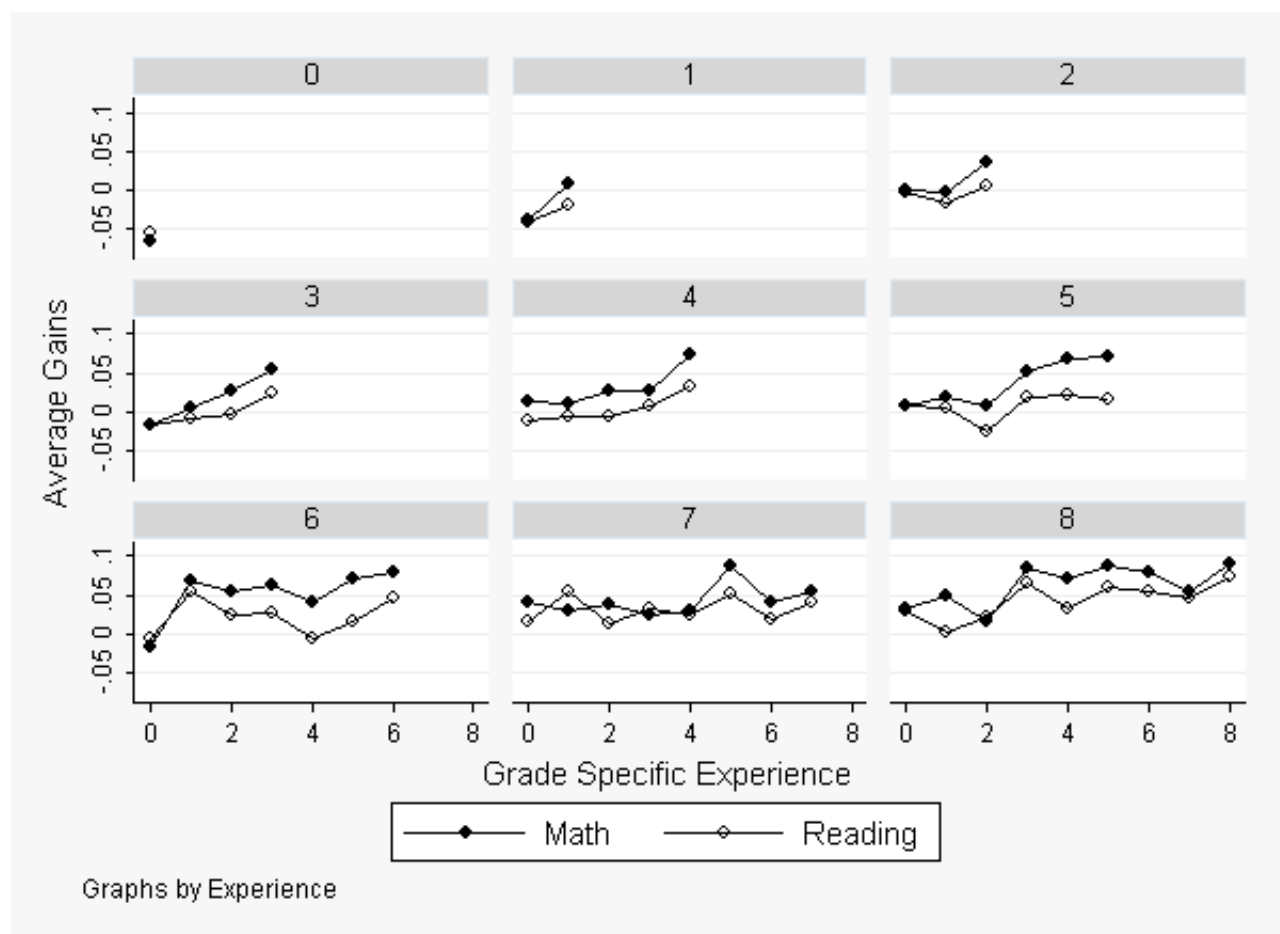
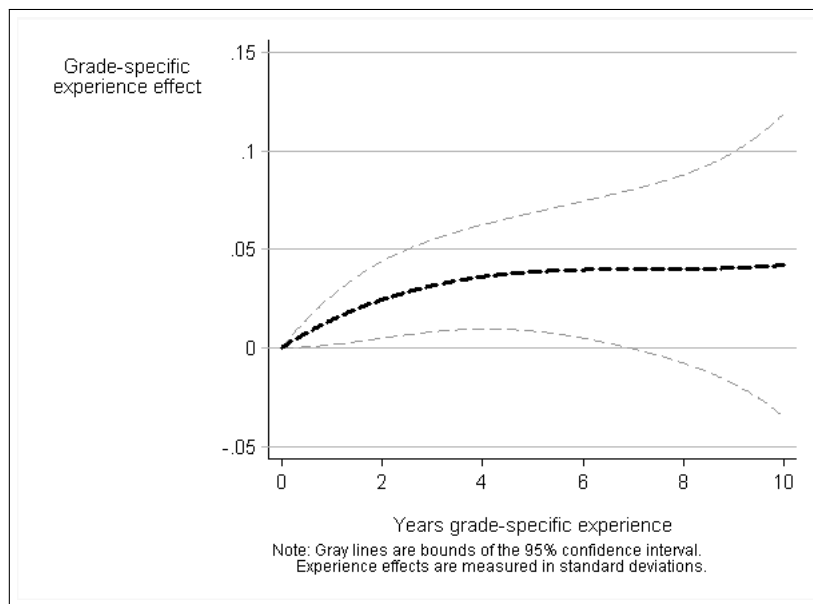


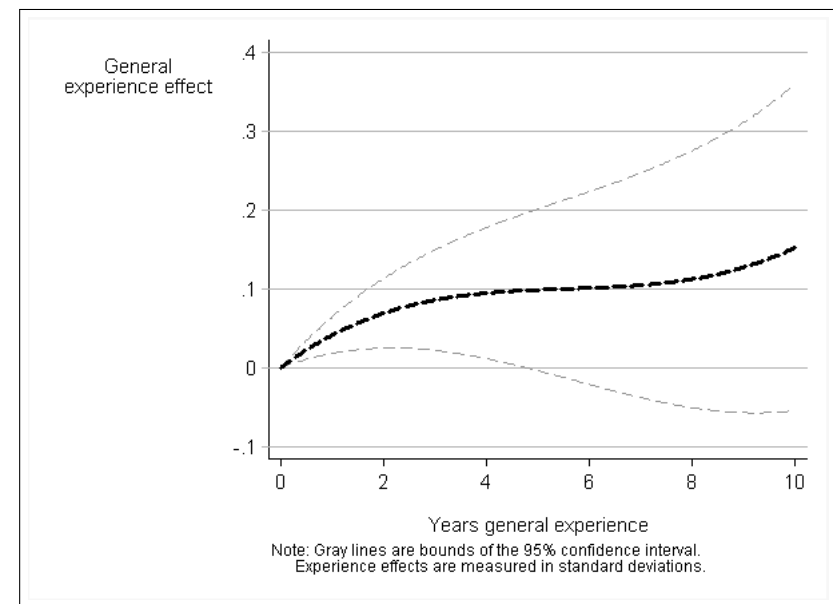
Figure 3: Average student test score gains by teacher grade specific experience: Split by experience level.

Notes: As experience rises, the number of teachers reflected in each point sharply decreases and thus the plots become increasingly noisy. For example, there are relatively few teachers with exactly 8 years of experience *and* exactly 6 years of grade-specific experience.

Figure 4: Return to Experience – Math Scores

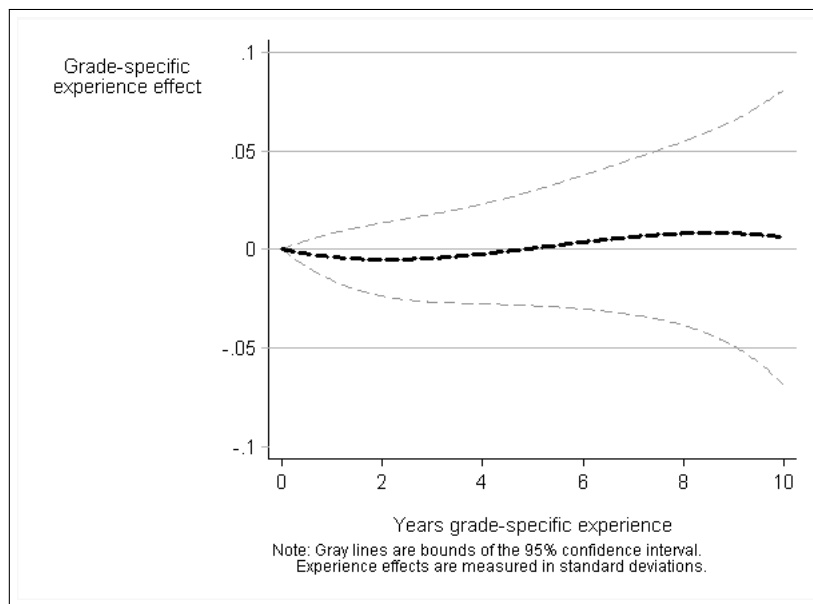


(a) Grade-specific Experience

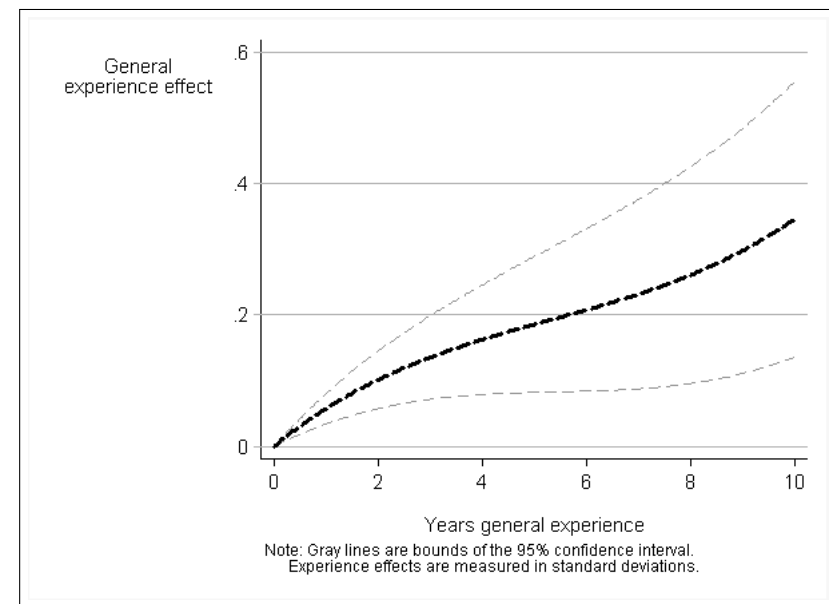


(b) General Experience

Figure 5: Return to Experience – Reading Scores



(a) Grade-specific Experience



(b) General Experience

Table 1: Descriptive Statistics

Variable	Full Sample			Restricted Sample		
	Observations	Mean	Standard deviation	Observations	Mean	Standard deviation
Unit of observation: Student-year						
Math Score	2188251	0.041	0.99	688962	-0.02	0.984
Reading Score	2178575	0.037	0.988	685667	-0.03	0.988
Change in math score	1840340	0.026	0.704	562480	0.011	0.733
Change in reading score	1832654	0.014	0.732	560052	-0.006	0.764
Female	2189917	0.495	0.5	689539	0.495	0.5
Black	2189917	0.284	0.451	689539	0.302	0.459
Hispanic	2189917	0.055	0.228	689539	0.068	0.252
Parent is high school dropout	2189917	0.094	0.291	689539	0.082	0.274
Parent is high school graduate	2189917	0.560	0.496	689539	0.542	0.498
Parent is college graduate	2189917	0.236	0.425	689539	0.221	0.415
Class size	2189917	22.909	3.313	689539	22.525	3.362
Student has limited english proficiency	2189917	0.019	0.137	689539	0.022	0.146
Unit of observation: Teacher-year						
Experience	100681	12.051	9.585	32306	2.911	2.715
Grade-specific experience (No Imputation)	-	-	-	32306	1.842	2.109
Grade-specific experience (Imputed)	100681	5.685	3.935	-	-	-
Female Teacher	100681	0.928	0.259	32306	0.878	0.327
Black Teacher	100681	0.138	0.345	32306	0.113	0.317
Hispanic Teacher	100681	0.003	0.059	32306	0.005	0.068
Teacher has advanced degree	100681	0.256	0.436	32306	0.131	0.338
Teacher Exam Score	93993	0.039	0.835	29913	0.152	0.702

Table 2: Grade Assignment Transition Matrix

Grade taught in year t	Grade taught in year t+1									Total
	PK	K	1	2	3	4	5	6	7 and up	
PK	80.8%	11.2%	2.9%	1.9%	1.2%	0.7%	0.4%	0.3%	0.6%	100.0%
K	1.2%	76.7%	10.4%	4.0%	2.6%	1.7%	1.2%	1.0%	1.2%	100.0%
1	0.4%	10.2%	72.3%	8.5%	3.5%	2.0%	1.4%	0.8%	0.8%	100.0%
2	0.3%	5.3%	8.8%	69.7%	9.2%	3.2%	2.1%	0.7%	0.7%	100.0%
3	0.2%	4.0%	4.5%	7.3%	69.6%	8.2%	3.7%	1.4%	1.1%	100.0%
4	0.2%	2.8%	2.9%	4.3%	7.8%	69.8%	8.9%	1.9%	1.5%	100.0%
5	0.2%	2.2%	2.1%	2.7%	5.4%	8.4%	72.1%	4.3%	2.6%	100.0%

Note: This table is restricted to teachers who teach in both year t and t+1 and are fully observed in the data.

Table 3: Grade Specific Experience by Total Experience

Experience	Grade Specific Experience						Total
	0	1	2	3	4	5	
0	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100%
1	32.9%	67.1%	0.0%	0.0%	0.0%	0.0%	100%
2	24.1%	23.3%	52.6%	0.0%	0.0%	0.0%	100%
3	20.2%	17.5%	19.2%	43.1%	0.0%	0.0%	100%
4	18.0%	14.5%	14.0%	16.6%	36.9%	0.0%	100%
5	16.1%	13.4%	11.9%	12.4%	15.6%	30.6%	100%

Note: This table includes all teachers in the fully observed sample.

Table 4: Test for Dynamic Endogeneity of Grade-Specific Experience

	Restricted sample				Full sample			
	Prob. teacher is switched between year t and t+1		Grade-specific exper. in year t+1		Prob. teacher is switched between year t and t+1		Grade-specific exper. in year t+1	
Average math score in year t	-0.013 (0.009)		0.019 (0.028)		-0.006 (0.004)		0.013 (0.015)	
Average read score in year t	0.000 (0.009)		-0.001 (0.028)		-0.004 (0.004)		-0.001 (0.014)	
Average math gains in year t	-0.005 (0.006)		0.005 (0.017)		-0.001 (0.002)		0.012 (0.009)	
Average read gains in year t	-0.009 [†] (0.005)		-0.009 (0.016)		-0.004* (0.002)		-0.010 (0.008)	
# of course offering changes between t and t+1	0.000 (0.002)	0.000 (0.002)			0.002* (0.001)	0.002* (0.001)		
Grade-specific exper. in year t	0.061** (0.008)	0.061** (0.007)	0.518** (0.034)	0.542** (0.034)	0.035** (0.003)	0.035** (0.002)	0.587** (0.014)	0.602** (0.013)
General experience in year t	-0.011 (0.025)	-0.012 (0.022)	0.387** (0.090)	0.399** (0.079)	0.004* (0.002)	0.003 [†] (0.002)	0.120** (0.007)	0.118** (0.006)
Teacher was switched between t and t+1			-1.931** (0.077)	-1.950** (0.078)			-2.102** (0.049)	-2.115** (0.048)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effect:	Teacher-by- School	Teacher-by- School	Teacher-by- School	Teacher-by- School	Teacher-by- School	Teacher-by- School	Teacher-by- School	Teacher-by- School
Observations	18808	20226	18974	20409	67036	71364	67630	71987

[†] Significant at 10%; * significant at 5%; ** significant at 1%. Standard errors clustered at the school-year are reported in parentheses

Notes: The observation is at the class level so test scores (and gains) are class averages. The results are nearly identical when estimated at the student level with standard errors clustered at the class level. For the first two columns, the dependent variable is an indicator for whether a teacher switches grades between year t and year t+1. For the third and fourth columns the dependent variable is grade-specific experience in year t+1. Columns five through eight repeat the tests using the entire sample (with grade-specific experience imputed).

Table 5: Impact of Teacher Experience on Student Performance

	Math			Reading		
	Teacher-by-School (1)	School (2)	Teacher-by-School-by-Grade (3)	Teacher-by-School (4)	School (5)	Teacher-by-School-by-Grade (6)
Fixed Effects:						
Grade-specific experience	0.0162* (0.0080)	0.0307** (0.0072)	0.0360* (0.0163)	-0.0054 (0.0076)	0.0041 (0.0062)	0.0129 (0.0159)
(Grade-specific experience) ²	-0.0022 (0.0021)	-0.0056** (0.0020)	-0.0029 (0.0023)	0.0016 (0.0021)	-0.0006 (0.0017)	0.0010 (0.0023)
(Grade-specific experience) ³	0.0001 (0.0002)	0.0002 (0.0001)	0.0001 (0.0002)	-0.0001 (0.0002)	0.0000 (0.0001)	-0.0001 (0.0002)
Experience	0.0492** (0.0126)	0.0386** (0.0067)	0.0346 [†] (0.0178)	0.0646** (0.0123)	0.0320** (0.0057)	0.0557** (0.0174)
Experience ²	-0.0084** (0.0018)	-0.0061** (0.0016)	-0.0080** (0.0020)	-0.0080** (0.0018)	-0.0052** (0.0014)	-0.0078** (0.0020)
Experience ³	0.0005** (0.0001)	0.0004** (0.0001)	0.0005** (0.0001)	0.0005** (0.0001)	0.0003** (0.0001)	0.0005** (0.0001)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Grade fixed effects	Yes	Yes	No	Yes	Yes	No
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Peer characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Observations	562480	562480	562480	560052	560052	560052

[†] Significant at 10%; * significant at 5%; ** significant at 1%. Standard errors clustered at class level reported in parentheses.

Notes: The dependent variable is test score. Only teachers who begin teaching during the sample frame are included in this regression. Appendix Table A-5 shows the complete set of coefficients from these regressions.

Table 6: Falsification: “Impact” of Future Teacher Experience on Current Student Performance

	Math			Reading		
	Teacher- by- School (1)	School (2)	Teacher-by- School-by- Grade (3)	Teacher- by- School (4)	School (5)	Teacher-by- School-by- Grade (6)
Fixed Effects:						
Lead teacher grade exper.	0.0045 (0.0046)	-0.0005 (0.0048)	0.0050 (0.0046)	0.0077 (0.0049)	0.0062 (0.0048)	0.0088 [†] (0.0049)
(Lead teacher grade exper.) ²	-0.0009 (0.0011)	0.0003 (0.0011)	-0.0010 (0.0011)	-0.0007 (0.0012)	-0.0006 (0.0011)	-0.0009 (0.0012)
(Lead teacher grade exper.) ³	0.0001 (0.0001)	-0.0000 (0.0001)	0.0001 (0.0001)	0.0000 (0.0001)	0.0000 (0.0001)	0.0000 (0.0001)
Lead teacher exper.	0.0072 (0.0058)	0.0087 (0.0059)	0.0077 (0.0057)	0.0010 (0.0062)	0.0000 (0.0059)	-0.0004 (0.0062)
(Lead teacher exper.) ²	-0.0019 (0.0013)	-0.0018 (0.0013)	-0.0020 (0.0013)	-0.0013 (0.0014)	-0.0008 (0.0014)	-0.0010 (0.0014)
(Lead teacher exper.) ³	0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)
Observations	261816	261816	261816	260743	260743	260743

[†] Significant at 10%; * significant at 5%; ** significant at 1%. Standard errors clustered at class level reported in parentheses.

Notes: The dependent variable is test score in period t. Lead experience and lead grade-specific experience correspond to the characteristics of student i's teacher in period t+1. All period t covariates are included in this regression.

Appendix

A Robustness Across Value Added Models

The lagged IV model used in the paper is my preferred specification both because it has performed well in experimental validation studies and because it can control for student sorting so long as that sorting is based on past test score. For robustness, I estimate other common VAMs in this appendix. Using the same variable definitions as presented in the paper, three commonly estimated VAM models are laid out below.

Gains model

$$\begin{aligned}\Delta A_{ijgst} = & \beta X_i + \delta C_{ijgst} + \rho V_{ij} + f(Exp_{jt}) + g(Expgrd_{jt}) \\ & + \pi D_{jt} + \xi_g + \omega_{js} + \phi_t + \epsilon_{ijgst}\end{aligned}\tag{6}$$

Lagged test score model

$$\begin{aligned}A_{ijgst} = & \alpha A_{ijt-1g_{t-1}s_{t-1},t-1} + \beta X_i + \delta C_{ijgst} + \rho V_{ij} \\ & + \pi D_{jt} + f(Exp_{jt}) + g(Expgrd_{jt}) + \xi_g + \omega_{js} + \phi_t + \epsilon_{ijgst}\end{aligned}\tag{7}$$

Student fixed-effect model

$$\begin{aligned}A_{ijgst} = & \beta X_i + \delta C_{ijgst} + \rho V_{ij} + f(Exp_{jt}) + g(Expgrd_{jt}) \\ & + \pi D_{jt} + \omega_{js} + \mu_i + \phi_t + \epsilon_{ijgst}\end{aligned}\tag{8}$$

The lagged test score model is identical to the preferred specification from the text except that it does not instrument for the lagged test score. This model assumes that all past inputs can be summarized by the lagged test score and it implicitly assumes that the effects of past inputs decay geometrically. The gains model additionally makes the assumption that $\alpha = 1$. One advantage of the gains model over the lagged model is that it completely avoids the measurement error problems of including the lagged score as an independent variable.

The student fixed effects model controls for unobserved fixed characteristics of students and assumes that past inputs have no effect on future outcomes. Unfortunately, given my data, the model with student fixed effects may be unreliable when estimated for the fully observed, restricted sample of teachers. This analysis is complicated by the fact that when I restrict the sample to fully observed teachers, this severely reduces the number of students who are observed for two or more years.²⁶ As a result, while restricting the sample to fully observed teachers is broadly my preferred specification, since it destroys the continuous panel structure along the student dimension, my preferred student fixed effect estimates are based on the full sample of teachers.

Even when using the full sample of students however, one drawback to using the student fixed effects model is that one cannot simultaneously control for grade fixed effects, year fixed effects and student fixed effects (see Rockoff (2004) for a discussion). Previous research has simply dropped the grade fixed effects; however, given that this paper's focus is differences between grades, the exclusion of grade fixed effects seems ill advised. Nevertheless, I report estimates from the student-fixed effects model (excluding grade fixed effects) with the understanding that the restricted sample estimates are identified off of very few students.

Estimates based on the above 3 models (as well as the preferred specification from the text) are shown in Table A-3. Results are extremely similar for the lagged model, the instrumented lagged model and the gains model. When including student fixed effects for the restricted sample, grade-specific experience appears to be much more important and the impact of general experience is not statistically significant. The differences in results for the student fixed effect model could be due to the exclusion of grade fixed effects or the controls for unobserved fixed student ability; however, I find it most likely that the differences are due to the lack of repetitive student observations because when estimated on the full sample of teachers, the student fixed effect model yields similar results to the other models. Table A-4 shows estimates of the same models for the full sample of teachers. When using the entire sample, all four value-added models yield qualitatively similar results.

²⁶If a student is observed for three years in the data, this student is only observed for three years in the restricted sample if he/she has a fully observed teacher in every year. Only 16% of the students who are observed for three years in the full sample, appear for three years in the restricted sample.

Table A-1: Impact of Teacher Experience on Student Performance - Full Sample

	Math			Reading		
	Teacher- by- School (1)	School (2)	Teacher-by- School-by- Grade (3)	Teacher- by- School (4)	School (5)	Teacher-by- School-by- Grade (6)
Fixed Effects:						
Grade-specific experience	0.0155** (0.0035)	0.0269** (0.0029)	0.0228** (0.0049)	0.0058 [†] (0.0032)	0.0140** (0.0024)	0.0096* (0.0045)
(Grade-specific experience) ²	-0.0020** (0.0007)	-0.0029** (0.0006)	-0.0028** (0.0008)	-0.0010 (0.0006)	-0.0017** (0.0005)	-0.0014 [†] (0.0007)
(Grade-specific experience) ³	0.0001 (0.0000)	0.0001** (0.0000)	0.0001* (0.0000)	0.0000 (0.0000)	0.0001** (0.0000)	0.0001 (0.0000)
Experience	0.0322** (0.0042)	0.0288** (0.0034)	0.0249** (0.0050)	0.0179** (0.0039)	0.0186** (0.0028)	0.0166** (0.0046)
Experience ²	-0.0069** (0.0008)	-0.0052** (0.0008)	-0.0061** (0.0009)	-0.0040** (0.0008)	-0.0032** (0.0006)	-0.0039** (0.0008)
Experience ³	0.0004** (0.0001)	0.0003** (0.0000)	0.0004** (0.0001)	0.0002** (0.0000)	0.0002** (0.0000)	0.0002** (0.0001)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Grade fixed effects	Yes	Yes	No	Yes	Yes	No
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Peer characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1840340	1840340	1840340	1832654	1832654	1832654

[†] Significant at 10%; * significant at 5%; ** significant at 1%. Standard errors clustered at class level reported in parentheses.

Notes: The dependent variable is test score. Teachers with out-of-sample experience are included and the distribution of grades taught out-of-sample is imputed as laid out in the text.

Table A-2: Falsification: “Impact” of Future Teacher Experience on Current Student Performance - Full sample

Fixed Effects:	Math			Reading		
	Teacher- by- School (1)	School (2)	Teacher-by- School-by- Grade (3)	Teacher- by- School (4)	School (5)	Teacher-by- School-by- Grade (6)
Lead teacher grade exper.	-0.0006 (0.0024)	-0.0004 (0.0026)	-0.0005 (0.0024)	0.0048 [†] (0.0026)	0.0030 (0.0026)	0.0050 [†] (0.0026)
(Lead teacher grade exper.) ²	0.0004 (0.0006)	0.0004 (0.0006)	0.0004 (0.0006)	-0.0008 (0.0006)	-0.0005 (0.0006)	-0.0008 (0.0006)
(Lead teacher grade exper.) ³	-0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
Lead teacher exper.	0.0062* (0.0030)	0.0062 [†] (0.0032)	0.0064* (0.0030)	-0.0007 (0.0032)	0.0004 (0.0032)	-0.0012 (0.0032)
(Lead teacher exper.) ²	-0.0014* (0.0007)	-0.0012 (0.0007)	-0.0015* (0.0007)	-0.0003 (0.0007)	-0.0002 (0.0007)	-0.0002 (0.0007)
(Lead teacher exper.) ³	0.0001* (0.0000)	0.0001 (0.0000)	0.0001* (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
Observations	901578	901578	901578	898129	898129	898129

[†] Significant at 10%; * significant at 5%; ** significant at 1%. Standard errors clustered at class level reported in parentheses.

Notes: The dependent variable is test score in period t. Lead experience and lead grade-specific experience correspond to the characteristics of student i’s teacher in period t+1. All period t covariates are included in this regression.

Table A-3: Robustness Across Value-Added Models

Model:	Math				Reading			
	Lagged IV (1)	Gains (2)	Lagged (3)	Student Fixed Effect (4)	Lagged IV (5)	Gains (6)	Lagged (7)	Student Fixed Effect (8)
Grade-specific experience	0.0162* (0.0080)	0.0180** (0.0066)	0.0165* (0.0076)	0.0462** (0.0075)	-0.0054 (0.0076)	-0.0041 (0.0060)	-0.0052 (0.0072)	0.0031 (0.0086)
(Grade-specific experience) ²	-0.0022 (0.0021)	-0.0030 [†] (0.0018)	-0.0023 (0.0020)	-0.0081** (0.0020)	0.0016 (0.0021)	0.0010 (0.0016)	0.0015 (0.0020)	0.0007 (0.0024)
(Grade-specific experience) ³	0.0001 (0.0002)	0.0001 (0.0001)	0.0001 (0.0002)	0.0005** (0.0002)	-0.0001 (0.0002)	-0.0001 (0.0001)	-0.0001 (0.0001)	0.0000 (0.0002)
Experience	0.0492** (0.0126)	0.0480** (0.0108)	0.0490** (0.0120)	0.0170 (0.0146)	0.0646** (0.0123)	0.0592** (0.0095)	0.0637** (0.0115)	0.0601** (0.0169)
Experience ²	-0.0084** (0.0018)	-0.0075** (0.0015)	-0.0083** (0.0017)	-0.0036* (0.0018)	-0.0080** (0.0018)	-0.0067** (0.0014)	-0.0078** (0.0017)	-0.0053* (0.0021)
Experience ³	0.0005** (0.0001)	0.0004** (0.0001)	0.0005** (0.0001)	0.0002 (0.0001)	0.0005** (0.0001)	0.0004** (0.0001)	0.0005** (0.0001)	0.0002 (0.0001)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade fixed effects	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Student characteristics	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Peer characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	562480	562480	562480	562480	560052	560052	560052	560052

[†] Significant at 10%; * significant at 5%; ** significant at 1%. Standard errors clustered at class level reported in parentheses.

Notes: The dependent variable is test score. Headings refer to model type. The Lagged IV model is the preferred specification used throughout the main text. The gains, lagged and student fixed effect models correspond to equations 6, 7, and 8 respectively

Table A-4: Robustness Across Value-Added Models - Full Sample

Model:	Math				Reading			
	Lagged IV (1)	Gains (2)	Lagged (3)	Student Fixed Effect (4)	Lagged IV (5)	Gains (6)	Lagged (7)	Student Fixed Effect (8)
Grade-specific experience	0.0155** (0.0035)	0.0145** (0.0031)	0.0153** (0.0034)	0.0246** (0.0021)	0.0058 [†] (0.0032)	0.0037 (0.0028)	0.0056 [†] (0.0032)	0.0137** (0.0024)
(Grade-specific experience) ²	-0.0020** (0.0007)	-0.0015* (0.0006)	-0.0019** (0.0007)	-0.0025** (0.0004)	-0.0010 (0.0006)	-0.0003 (0.0005)	-0.0009 (0.0006)	-0.0009 [†] (0.0005)
(Grade-specific experience) ³	0.0001 (0.0000)	0.0000 (0.0000)	0.0001 (0.0000)	0.0001** (0.0000)	0.0000 (0.0000)	-0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
Experience	0.0322** (0.0042)	0.0368** (0.0039)	0.0329** (0.0041)	0.0358** (0.0028)	0.0179** (0.0039)	0.0227** (0.0034)	0.0185** (0.0038)	0.0232** (0.0031)
Experience ²	-0.0069** (0.0008)	-0.0070** (0.0008)	-0.0069** (0.0008)	-0.0071** (0.0005)	-0.0040** (0.0008)	-0.0042** (0.0007)	-0.0040** (0.0008)	-0.0045** (0.0006)
Experience ³	0.0004** (0.0001)	0.0004** (0.0000)	0.0004** (0.0001)	0.0004** (0.0000)	0.0002** (0.0000)	0.0002** (0.0000)	0.0002** (0.0000)	0.0002** (0.0000)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade fixed effects	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Student characteristics	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Peer characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1840340	1840340	1840340	1840340	1832654	1832654	1832654	1832654

[†] Significant at 10%; * significant at 5%; ** significant at 1%. Standard errors clustered at class level reported in parentheses.

Notes: The dependent variable is test score. Headings refer to model type. The Lagged IV model is the preferred specification used throughout the main text. The gains, lagged and student fixed effect models correspond to equations 6, 7, and 8 respectively

Table A-5: Full Set of Coefficients from Table 5

	Math			Reading		
	School (1)	Teacher- by- School (2)	Teacher-by- School-by- Grade (3)	School (4)	Teacher- by- School (5)	Teacher-by- School-by- Grade (6)
Fixed Effects:						
Female student	-0.0013 (0.0032)	-0.0009 (0.0031)	-0.0015 (0.0031)	-0.0042 (0.0033)	-0.0034 (0.0033)	-0.0036 (0.0034)
Black student	-0.0198** (0.0032)	-0.0206** (0.0032)	-0.0197** (0.0032)	-0.0438** (0.0032)	-0.0386** (0.0034)	-0.0378** (0.0034)
Hispanic student	0.0183* (0.0077)	0.0223** (0.0077)	0.0233** (0.0077)	-0.0059 (0.0076)	-0.0039 (0.0076)	-0.0025 (0.0077)
Parental education						
High school dropout	-0.0138** (0.0044)	-0.0232** (0.0044)	-0.0244** (0.0044)	-0.0332** (0.0048)	-0.0400** (0.0048)	-0.0410** (0.0048)
College graduate	0.0086** (0.0032)	0.0179** (0.0032)	0.0190** (0.0032)	0.0014 (0.0033)	0.0098** (0.0033)	0.0107** (0.0034)
Student on subsidized lunch	-0.0297** (0.0026)	-0.0288** (0.0026)	-0.0285** (0.0026)	-0.0262** (0.0028)	-0.0259** (0.0028)	-0.0260** (0.0028)
Limited english proficiency	0.1234** (0.0085)	0.1214** (0.0085)	0.1207** (0.0085)	0.1539** (0.0093)	0.1522** (0.0093)	0.1535** (0.0093)
Student-teacher same sex	0.0051 [†] (0.0031)	0.0046 (0.0030)	0.0049 (0.0030)	-0.0073* (0.0033)	-0.0070* (0.0033)	-0.0069* (0.0033)
Student-teacher same ethnicity	-0.0052 (0.0070)	0.0018 (0.0087)	0.0007 (0.0087)	0.0113 [†] (0.0064)	-0.0093 (0.0092)	-0.0101 (0.0092)
Class size	-0.0032** (0.0006)	-0.0051** (0.0006)	-0.0050** (0.0007)	-0.0017** (0.0005)	-0.0035** (0.0006)	-0.0037** (0.0007)
Teacher was switched this year	0.0200** (0.0076)	0.0135 [†] (0.0078)	0.0173* (0.0083)	0.0009 (0.0065)	0.0003 (0.0076)	0.0048 (0.0081)
Total number of times switched	-0.0186** (0.0030)	0.0003 (0.0063)	-0.0014 (0.0106)	-0.0100** (0.0026)	-0.0057 (0.0061)	-0.0011 (0.0102)
Experience	0.0386** (0.0067)	0.0492** (0.0126)	0.0346 [†] (0.0178)	0.0320** (0.0057)	0.0646** (0.0123)	0.0557** (0.0174)
Experience ²	-0.0061** (0.0016)	-0.0084** (0.0018)	-0.0080** (0.0020)	-0.0052** (0.0014)	-0.0080** (0.0018)	-0.0078** (0.0020)
Experience ³	0.0004** (0.0001)	0.0005** (0.0001)	0.0005** (0.0001)	0.0003** (0.0001)	0.0005** (0.0001)	0.0005** (0.0001)

Continued on next page

Table A-5: Continued from previous page

	Math			Reading		
	School	Teacher-by-School	Teacher-by-School-by-Grade	School	Teacher-by-School	Teacher-by-School-by-Grade
Fixed Effects:						
Grade-specific experience	0.0307** (0.0072)	0.0162* (0.0080)	0.0360* (0.0163)	0.0041 (0.0062)	-0.0054 (0.0076)	0.0129 (0.0159)
(Grade-specific experience) ²	-0.0056** (0.0020)	-0.0022 (0.0021)	-0.0029 (0.0023)	-0.0006 (0.0017)	0.0016 (0.0021)	0.0010 (0.0023)
(Grade-specific experience) ³	0.0002 (0.0001)	0.0001 (0.0002)	0.0001 (0.0002)	0.0000 (0.0001)	-0.0001 (0.0002)	-0.0001 (0.0002)
Peer Characteristics						
Ave. Lagged math score	-0.2581** (0.0080)	-0.0314* (0.0124)	0.0243 [†] (0.0138)			
Ave. Lagged read score				-0.1575** (0.0067)	0.1057** (0.0115)	0.1671** (0.0128)
Fraction female	0.0320 [†] (0.0178)	0.0433* (0.0183)	0.0343 [†] (0.0191)	0.0211 (0.0154)	-0.0133 (0.0182)	-0.0251 (0.0191)
Fraction black	-0.2284** (0.0200)	-0.0332 (0.0222)	0.0043 (0.0233)	-0.1066** (0.0171)	0.0936** (0.0218)	0.1292** (0.0230)
Fraction hispanic	-0.2102** (0.0320)	-0.0827* (0.0380)	-0.0594 (0.0401)	-0.1988** (0.0295)	-0.1070** (0.0380)	-0.0730 [†] (0.0401)
Frac. parent ed: Dropout	0.0494** (0.0170)	-0.0149 (0.0199)	-0.0179 (0.0208)	0.0418** (0.0149)	0.0253 (0.0199)	0.0264 (0.0209)
Frac. parent ed: College grad.	0.0102 (0.0124)	0.0121 (0.0146)	0.0060 (0.0154)	-0.0058 (0.0110)	-0.0106 (0.0142)	-0.0183 (0.0150)
Frac. subsidized lunch	-0.0832** (0.0109)	0.0098 (0.0117)	0.0266* (0.0124)	-0.0501** (0.0094)	0.0425** (0.0116)	0.0520** (0.0124)
Frac. limited english proficient	0.0738 [†] (0.0391)	0.1625** (0.0438)	0.1564** (0.0459)	0.0709 [†] (0.0362)	0.2479** (0.0444)	0.2920** (0.0471)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Grade fixed effects	Yes	Yes	No	Yes	Yes	No
Observations	562480	562480	562480	560052	560052	560052

[†] Significant at 10%; * significant at 5%; ** significant at 1%. Clustered standard errors reported in parentheses

Notes: The dependent variable is test score. This Table provides the full set of coefficients for the regressions presented in Table 5