American Economic Journal: Applied Economics 2014, 6(2): 127–151
http://dx.doi.org/10.1257/app.6.2.127

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

By Ben Ost*

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 benefits of general teaching experience and specific curriculum familiarity. I find that both specific and general human capital contribute to teacher improvement and that recent specific experience is more valuable than distant specific experience. This paper also contributes to a broader literature on human capital acquisition, as it is among the first to examine human capital specificity using a direct measure of productivity. (*JEL* H75, I21, J24, J45)

The degree to which human capital acquisition is general or specific has been of central concern in the labor economics literature since Becker (1964). While this literature has made great strides in understanding the degree to which human capital is transferable across industries, firms, occupations and tasks, most studies measure productivity implicitly by assuming that wages perfectly reflect productivity in every time period. This assumption fails to hold in the presence of wage-deferring contracts, differential monopsony power, or efficiency wages and thus, any of these phenomena could bias estimates of the degree of human capital specificity. Importantly, these phenomena can generate the appearance of human capital specificity even when human capital is entirely general. There is little direct productivity-based evidence regarding human capital specificity because few datasets include the repeated productivity measures necessary to implement such an analysis.

This paper provides new evidence of the relative importance of general and task-specific human capital by examining productivity improvement among teachers. A large literature has established that teachers improve with experience, but no previous study on teacher improvement has made the distinction between general teaching human capital and human capital that is specific to a particular grade level. Using micro-level longitudinal data, I track teachers, their grade-level assignments

* University of Illinois at Chicago, 601 S. Morgan UH 718 Chicago, IL 60607 (e-mail: bost@uic.edu). I would like to thank Marcus Casey, James Cowan, Ron Ehrenberg, Kirabo Jackson, Robert Kaestner, Darren Lubotsky, Jordan Matsudaira, Mark Montgomery, Carrie Ost, Javaeria Qureshi, Steve Rivkin, Jonah Rockoff, Houston Stokes, Doug Webber, Ken Whelan, and anonymous reviewers for very helpful comments on earlier drafts. I thank Jesse Rothstein for providing some of the code used to process the teacher data. Also, I am indebted to the North Carolina Education Research Data Center for making available the confidential data used in this study.

† Go to http://dx.doi.org/10.1257/app.6.2.127 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

and a direct measure of productivity over an 18-year period. Using these data, I estimate the productivity improvements made by teachers as they gain general and grade-specific experience. This analysis provides estimates of how the entire history of teacher task assignments interact to determine current productivity.[1]

The literature on teacher improvement has developed a high level of rigor in recent years thanks to the availability of matched teacher-student panel data. Unlike the vast majority of data used to examine worker improvement, the data on teacher improvement include annual measures of productivity and are thus uniquely appropriate to directly examine a variety of theories regarding productivity growth. Furthermore, rather than relying on cross-sectional estimates that are likely flawed due to survival bias, researchers have used within-teacher variation to determine productivity improvements due to experience (Rockoff 2004; Hanushek et al. 2005; Aaronson, Barrow, and Sander 2007; Clotfelter, Ladd, and Vigdor 2007).

This paper is the first to document two stylized facts. First, I show that teachers switch grade assignments frequently within a school such that less than half remain in the same grade in their first five years. Second, students who have a teacher with more grade-specific experience make larger test score improvements than students who have a similarly experienced teacher with less grade-specific experience. The fact that teachers are regularly switched is critical to the implementation of my analysis because it suggests that sufficient variation exists between general and grade-specific experience. While suggestive, the higher value-added of teachers with more grade-specific experience should not be taken as conclusive evidence that grade-specific experience is beneficial to teacher productivity. The difference in productivity between teachers of equal general experience levels could be the result of teachers improving with specific experience, but could also potentially reflect differences in grade assignment patterns across different types of schools and teachers.

My analysis distinguishes between these possibilities by carefully considering the source of identifying variation and implementing tests for endogenous movement. First, because of the possibility that unobserved teacher characteristics are correlated with grade-specific experience, my preferred specification controls for teacher fixed effects and thus the primary findings are based on comparing a teacher to herself. Second, in order to test for endogenous movement, I examine whether changes in grade-specific experience are predicted by current performance. Furthermore, I test whether the type of students a teacher is assigned is related to the teacher's grade-specific experience.

My preferred specification, which includes teacher fixed effects, finds that a teacher performs relatively better when she has more years of grade-specific experience, holding constant her total years of experience. As measured by her students' math score improvements, grade-specific experience is found to be approximately one-third to one-half as important as general teaching experience. For reading scores,

[1] Value-added estimates may be limited because it measures only teacher contributions towards instantaneous test score growth and therefore fails to capture other important aspects of a teacher's job. That said, Chetty, Friedman, and Rockoff (2011) find that a teacher's quality as measured by value-added is also reflected in her students' adult outcomes.

grade-specific experience does not appear to contribute to student improvement. One potential explanation for why repeating grade assignments benefits math but not reading scores is that in North Carolina the reading objectives are constant across grades whereas the math objectives change each year.

Beyond assessing the direct effect of grade-specific experience on student test scores, I explore the extent to which grade-specific human capital depreciates over time. While grade-specific experience contributes to teacher productivity, I find that the timing of this experience is important. Teachers who have recently taught their current grade derive substantial benefit from that experience, but distant grade-specific experience provides no added benefit. Using variation in the timing of grade-specific human capital acquisition, I estimate a nonlinear model that includes a depreciation parameter that discounts grade-specific experience according to when it was acquired. The depreciation parameter is estimated directly from the model and suggests that approximately 35 percent of specific human capital depreciates each year.

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 one of the first productivity-based estimates of how specific human capital depreciates over time. Lastly, the results of this paper provide a more nuanced understanding of how teachers improve and can guide policy regarding teacher grade assignments and professional development.

I. Literature

While the impact of many teacher characteristics is still debated, there exists an emerging consensus that teacher experience positively contributes to student learning, particularly for younger grades. Using data on middle school and elementary school students in Texas, Hanushek et al. (2005) find that students perform relatively worse when their teacher has less than three years of experience. Rockoff (2004) finds consistent results using matched teacher-student data from two New Jersey elementary school districts. Similarly, Clotfelter, Ladd, and Vigdor (2007) and Jackson and Bruegmann (2009) use the same North Carolina matched teacher-student data used in this paper and find that elementary teachers improve with experience, especially in the first several years.

The one exception to this consensus is Aaronson, Barrow, and Sander (2007). The authors use data for ninth graders in the Chicago Public Schools and find no evidence of teachers improving with experience. One potential explanation for why the Aaronson, Barrow, and Sander (2007) results differ from other studies is that most previous research has focused on students in grades 3–8 whereas Aaronson, Barrow, and Sander (2007) considers high school teachers. For elementary grades, the fact that teachers typically teach the same students all day makes it more likely that differences in teaching ability will be detectable through student performance. Second, it is possible that the key skills that teachers develop as they gain experience are useful for teaching younger students but not for secondary education. Aaronson, Barrow, and Sander (2007) is a clear exception to the literature given that in a meta-analysis of the value-added literature, Harris (2009) finds that eight of nine studies show evidence of teachers improving with experience.

While many papers have demonstrated that teachers improve, the only paper I am aware of that explores a mechanism for *how* teachers' on-the-job experience helps them 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.

II. Data

I use longitudinal administrative data that link students to their teachers in the state of North Carolina between 1995–2012.[2] These data include 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, limited English status, 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.[3] 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 school teachers often teach multiple grades simultaneously, and because student test score data are most complete for third through fifth grade, I restrict my sample to elementary teachers who teach third through fifth grade single-grade classes.

While the North Carolina data include a link between student test scores and teachers, until 2006, the teacher listed is actually the proctor of the student exam, and 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, Ladd, and Vigdor (2007); Rothstein (2010); 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 and the classroom has more than ten students. In addition, I drop cases where a proctor administered more than half of his/her tests to a different grade level.[4]

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

[2] These data have been extensively cleaned and standardized by the North Carolina Education Research Data Center housed at Duke University.

[3] These data are described in great detail in Clotfelter, Ladd, and Vigdor (2007).

[4] In describing these 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."

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 include complete teacher information starting from 1995, complete student data are only available starting in 1997. I use the entire 1995–2012 period to calculate grade-specific experience, but only use 1997–2012 in the regressions.

As discussed by Koedel and Betts (2010), achievement tests that contain ceilings may lead to systematic measurement error since students near the ceiling are unable to make further gains. In results not shown, I test for a ceiling in the North Carolina data by comparing a kernel density of each distribution to that of the normal density and find no cause for concern.

A. *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 provide no information regarding the grades 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. In order to accurately calculate grade-specific experience, my preferred specification restricts the sample to fully observed teachers, though results are robust to including the entire sample and imputing grade-specific experience. The fact that the resulting 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, Hanushek, and Kain 2005).

Table 1 shows descriptive statistics for both the full sample and the sample that 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.[5] In addition, 27 percent of the full sample of teachers have an advanced degree whereas only 12.5 percent 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 the 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.

[5] By definition, teachers in the restricted sample must have less than 18 years of experience, whereas the full sample includes many teachers with over 30 years of experience.

Table 1—Descriptive Statistics

	Full sample			Restricted sample		
Variable	Observations	Mean	SD	Observations	Mean	SD
Unit of observation: Student-year						
Math score	3,681,863	0.024	0.997	656,441	−0.066	0.986
Reading score	3,668,021	0.02	0.996	653,851	−0.075	0.993
Change in math score	3,109,559	0.00	0.604	548,703	−0.016	0.606
Change in reading score	3,106,922	−0.01	0.637	549,299	−0.031	0.641
Female	3,696,271	0.496	0.5	659,265	0.497	0.5
Black	3,696,178	0.274	0.446	659,261	0.301	0.459
Hispanic	3,696,178	0.077	0.266	659,261	0.096	0.294
Class size	3,790,521	22.349	4.18	676,451	22.014	3.929
Student has limited English proficiency	3,671,838	0.048	0.213	654,226	0.061	0.24
Free or reduced lunch	3,675,428	0.546	0.498	659,922	0.593	0.491
Unit of observation: Teacher-year						
Experience	183,928	12.097	9.492	33,097	2.592	2.892
Grade-specific experience	—	—	—	31,075	1.786	2.27
Female teacher	183,928	0.924	0.265	33,097	0.877	0.328
Black teacher	183,928	0.133	0.34	33,097	0.104	0.305
Hispanic teacher	183,928	0.004	0.062	33,097	0.006	0.076
Teacher has advanced degree	170,344	0.27	0.444	30,830	0.125	0.331

III. General Experience versus 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 20 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. Evaluating 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.

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 for teachers in their first school. This table is restricted to teachers who have not switched schools to demonstrate that the divergence between general and grade-specific experience is not driven by school switching. Approximately 18 percent of teachers teach a new grade in their second year of teaching, and less than half 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 and general experience for this sample.

In light of the fact that this paper shows that grade switching has some negative effects, it is somewhat surprising that so many switches occur. Conversations with principals indicate that in addition to the costs of grade switching documented in this paper, several benefits to grade switching exist which may explain why teachers

Table 2—Grade Assignment Transition Matrix

Grade taught in year t	Grade taught in year $t+1$										
	PK	K	1	2	3	4	5	6	7	8 and up	Total
PK	93.3	2.9	1.6	0.9	0.5	0.4	0.2	0.2	0.1	0.0	100.0
KG	0.8	84.1	8.1	3.5	1.7	1.0	0.6	0.3	0.1	0.0	100.0
1	0.2	2.7	85.6	6.6	2.6	1.4	0.8	0.3	0.0	0.0	100.0
2	0.1	1.3	5.0	81.7	7.4	2.5	1.5	0.4	0.1	0.0	100.0
3	0.1	0.8	2.5	5.7	80.7	6.2	3.3	0.7	0.1	0.1	100.0
4	0.1	0.6	1.7	3.0	5.1	80.6	7.3	1.2	0.2	0.1	100.0
5	0.1	0.4	1.0	2.2	4.1	6.1	83.1	2.5	0.4	0.2	100.0

Notes: This table is restricted to teachers whose tenure is fully observed in the data, but it includes teachers who are not matched to student test scores. The table includes 125,312 teacher-year observations. The number of observations in this table is much larger than that used in the rest of the paper, because it includes teachers from all grade levels.

Table 3—Grade Specific Experience by Total Experience

Experience	Grade specific experience						
	0	1	2	3	4	5	Total
0	100.0	0.0	0.0	0.0	0.0	0.0	100.0
1	17.7	82.3	0.0	0.0	0.0	0.0	100.0
2	13.8	16.0	70.3	0.0	0.0	0.0	100.0
3	13.2	11.6	14.8	60.4	0.0	0.0	100.0
4	13.7	10.0	10.4	13.5	52.4	0.0	100.0
5	11.9	10.8	8.1	10.1	13.2	45.8	100.0

Notes: The experience variable is the total number of years teaching. Grade-specific experience is the number of prior years having taught the grade a teacher is currently teaching. Unlike in Table 2, this table only shows switching patterns for the first six years of teaching. The table includes 25,745 teacher-year observations. Rows do not add up to exactly 100 percent due to rounding.

switch grade assignments so frequently. First, switching teachers allows for a flexibility in management that can be useful when teacher teams conflict, lack diversity, or are uniformly experienced or inexperienced. Second, teachers who intend to become administrators may eventually benefit from the breadth of experience which comes from teaching a variety of courses. Third, it is possible that teachers grow bored with repetition over time and although they are better able to improve student test scores, their enthusiasm for the subject may wane. Finally, several principals indicated that they switch teachers to facilitate "professional growth history"—helping teachers become well rounded by having them teach a variety of grades.

No empirical estimates of the benefits to switching are available from the literature since no previous research has considered either the benefits or costs of teacher movement across grades within a school. Although this paper presents evidence that switching grades can be disruptive, the policy decision of whether to switch teachers must consider both the costs and benefits of switching. A complete cost-benefit analysis is beyond the scope of this paper; however, in the Identification Tests section, I examine whether patterns of teacher switching may be systematic in a way that would bias estimates of the return to specific experience.

A. *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 1, panel A and Figure 1, panel B graphically show 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. These figures make no sample restrictions and include no controls so they reflect the pattern in it's rawest form.

While these figures are suggestive, they simply reflect raw correlations and by themselves cannot be interpreted as implying a causal relationship. To test this relationship more rigorously, I place the analysis in a regression context.

IV. 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, results are robust to the use of other value-added models.[6]

$$(1) \quad A_{ijgst} = \alpha A_{i,t-1} + \beta \mathbf{X}_i + \delta \mathbf{C}_{ijgst} + \rho \mathbf{V}_{ij} + f(Exp_{jt}) + g(Expgrd_{jt}) + \xi_{gt} + \omega_j + \varepsilon_{ijgst}.$$

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 $\mathbf{X}_i$ includes student gender, ethnicity, subsidized lunch status, and parental education. Classroom characteristics such as class size are denoted by $\mathbf{C}_{ijgst}$. The vector $\mathbf{V}_{ij}$ includes interactions between the student and teacher ethnicity and sex.[7] This model includes grade-by-year fixed effects denoted by ξ_{gt}. Following the methodology of Papay and Kraft (2010), I estimate grade-by-year fixed effects by estimating equation (1), omitting the teacher fixed effects. I then use these estimated grade-by-year fixed effects to estimate equation (1) with the inclusion of teacher fixed effects.[8] Experience and grade-specific experience enter through $f(\cdot)$ and $g(\cdot)$, which

[6] The lagged test score VAM is used in many recent studies including Aaronson, Barrow, and Sander (2007); Kane, Riegg, and Staiger (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).

[7] Dee (2005) shows that gender and ethnicity match may affect student achievement.

[8] See Papay and Kraft (2010) for a detailed discussion of the advantage of estimating the value-added model in this way as opposed to simply including the grade-by-year fixed effects directly. Because the grade-by-year effects aim to capture variation in factors such as test difficulty, I estimate the grade-by-year effects using the unrestricted sample of students. Despite the methodological advantages that Papay and Kraft (2010) discuss, the results presented in this paper are very similar regardless of how the grade-by-year effects are estimated.

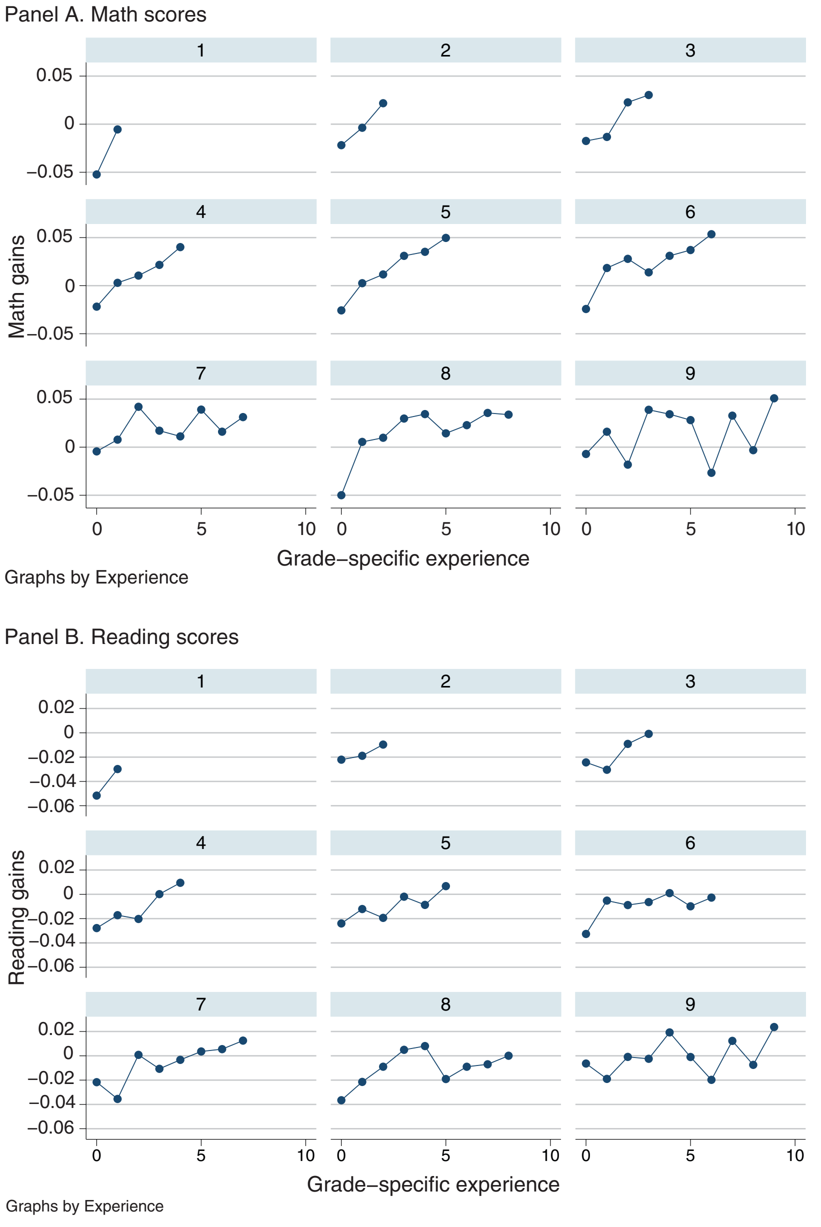

Figure 1. Average Student Gains by Teacher Grade-Specific Experience: Split by Experience Level

Notes: Each graph in Figure 1, panel A and Figure 1, panel B holds experience constant and shows the average test score gains for teachers with different levels of grade-specific experience. For example, the right most dot on the upper left graph of Figure 1, panel A shows the average math score gains made by students who were taught by a teacher with one-year of general experience and one-year of specific experience. This corresponds to a teacher who is in her second year of teaching and is teaching the same course as in her first year.

I model flexibly as a series of dummy variables. The teacher fixed effect is denoted by ω_j and in some specifications, I replace this with a school fixed effect. Unless otherwise specified, all specifications cluster standard errors at the classroom level. To avoid confounding grade-specific experience with school-specific experience, I restrict the analysis to teachers who have not switched schools.

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 (1) for the entire sample.[9]

While Rothstein (2010) demonstrates significant nonrandom 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 identification section and find little evidence that within a teacher, students are differentially sorted as the teacher gains grade-specific experience.

V. Results

Results from estimating equation (1) are shown in Table 4. Consistent with previous research, a teacher's experience is found to positively impact student outcomes. While controlling for the number of years of teaching (general experience), grade-specific experience has a positive impact on student math scores, but no effect on reading scores.

Column 2 of Table 4 shows that teachers with grade-specific experience outperform teachers with no experience teaching the grade.[10] The grade-specific experience effect is approximately half as large as the general experience effect, though this varies somewhat across the experience profile. These results highlight the importance of modeling experience and grade-specific experience in a flexible fashion since the dummy coefficients demonstrate that experience effects are highly nonlinear. These magnitudes show fairly substantial improvement relative to the overall distribution of teacher quality and are consistent with previous estimates

[9] See the online Appendix of Jackson and Bruegmann (2009) for a proof of the consistency of this estimator. In practice, using this estimator, a simple gains model or a lagged IV model all yield similar estimates as shown in Table A1.

[10] For the remainder of the manuscript, I refer to this specification as the baseline specification.

TABLE 4—IMPACT OF TEACHER EXPERIENCE ON STUDENT PERFORMANCE

	Math				Reading			
Fixed effects:	Teacher (1)	Teacher (2)	School (3)	School (4)	Teacher (5)	Teacher (6)	School (7)	School (8)
Exp=1	0.0580***	0.0470***	0.0755***	0.0606***	0.0283***	0.0321***	0.0370***	0.0339***
	(0.0034)	(0.0059)	(0.0043)	(0.0068)	(0.0031)	(0.0053)	(0.0034)	(0.0054)
Exp=2	0.0785***	0.0555***	0.1007***	0.0725***	0.0449***	0.0464***	0.0602***	0.0560***
	(0.0041)	(0.0070)	(0.0047)	(0.0072)	(0.0037)	(0.0064)	(0.0038)	(0.0059)
Exp=3	0.0886***	0.0583***	0.1110***	0.0793***	0.0453***	0.0381***	0.0608***	0.0461***
	(0.0047)	(0.0082)	(0.0052)	(0.0078)	(0.0043)	(0.0074)	(0.0043)	(0.0064)
Exp=4	0.0895***	0.0528***	0.1175***	0.0780***	0.0523***	0.0467***	0.0753***	0.0606***
	(0.0055)	(0.0092)	(0.0060)	(0.0087)	(0.0050)	(0.0083)	(0.0050)	(0.0071)
Exp=5	0.0903***	0.0587***	0.1210***	0.0826***	0.0460***	0.0429***	0.0691***	0.0519***
	(0.0063)	(0.0105)	(0.0068)	(0.0097)	(0.0057)	(0.0092)	(0.0055)	(0.0077)
Exp=6	0.0943***	0.0576***	0.1275***	0.0789***	0.0395***	0.0394***	0.0633***	0.0429***
	(0.0073)	(0.0114)	(0.0080)	(0.0105)	(0.0064)	(0.0102)	(0.0061)	(0.0086)
Exp=7	0.0907***	0.0626***	0.1318***	0.0842***	0.0461***	0.0490***	0.0743***	0.0582***
	(0.0085)	(0.0132)	(0.0093)	(0.0123)	(0.0079)	(0.0121)	(0.0077)	(0.0101)
Exp $\geq$ 8	0.0983***	0.0766***	0.1411***	0.0934***	0.0494***	0.0580***	0.0834***	0.0663***
	(0.0073)	(0.0138)	(0.0067)	(0.0106)	(0.0065)	(0.0120)	(0.0054)	(0.0085)
Expgrd=1		0.0133**		0.0198***		−0.0054		0.0033
		(0.0053)		(0.0062)		(0.0049)		(0.0050)
Expgrd=2		0.0287***		0.0384***		−0.0007		0.0069
		(0.0066)		(0.0069)		(0.0059)		(0.0056)
Expgrd=3		0.0350***		0.0406***		0.0099		0.0218***
		(0.0079)		(0.0079)		(0.0071)		(0.0065)
Expgrd=4		0.0418***		0.0521***		0.0063		0.0206***
		(0.0092)		(0.0093)		(0.0083)		(0.0075)
Expgrd=5		0.0318***		0.0474***		0.0016		0.0230***
		(0.0109)		(0.0108)		(0.0099)		(0.0088)
Expgrd=6		0.0440***		0.0679***		−0.0018		0.0302***
		(0.0127)		(0.0126)		(0.0112)		(0.0103)
Expgrd=7		0.0183		0.0611***		−0.0119		0.0162
		(0.0151)		(0.0149)		(0.0135)		(0.0121)
Expgrd $\geq$ 8		0.0116		0.0559***		−0.0131		0.0239**
		(0.0158)		(0.0130)		(0.0135)		(0.0104)
Student covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Peer characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher covariates	No	No	Yes	Yes	No	No	Yes	Yes
Year-by-grade FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	538,604	524,413	529,945	515,936	514,500	500,276	507,244	493,229

Notes: The dependent variable is a standardized measure of test score. Only teachers who begin teaching during the sample frame are included in this regression. Each line in the table corresponds to a separate dummy variable that is unity for a particular experience/grade-experience level, and zero otherwise. Standard errors clustered at class level reported in parentheses.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

of the return to experience. To put these magnitudes in perspective, the difference between a third year teacher teaching a new grade and a third year teacher who has always taught the same grade is approximately one-third as large as the impact of moving to a high-performing charter school (Hoxby and Murarka 2009).

Column 6 of Table 4 shows the same model estimated for reading score gains. Unlike for math scores, there is no evidence that grade-specific human capital is important for reading score improvement. While I have no definitive explanation as to why grade-specific experience matters more for math than for reading, one possible cause is the fact that similar reading skills are taught in each grade, whereas math curricula change dramatically for each grade. The 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, while all five math competency goals change for each grade (North Carolina Department of Education 2009).

The fourth and eighth columns of Table 4 show that results are similar when using school fixed effects, but the grade-specific experience effects are larger in magnitude. The fact that part of the grade-specific experience effect can be explained by unobserved heterogeneity in teacher quality suggests that more able teachers accrue grade-specific experience more rapidly than less able teachers. That said, the qualitative results do not depend on the level of fixed effect and for most of the indicators, the school fixed effect and teacher fixed effect specifications are statistically indistinguishable. In any case, to avoid the possibility that differential sorting or attrition bias estimates, my preferred specification includes teacher fixed effects.

A. *Grade-Specific Human Capital Depreciation*

The specification shown in Table 4 assumes that specific human capital does not decay over time—grade-specific human capital accrued last year is treated as equivalent to grade-specific human capital accrued five years earlier. To explore whether recent grade-specific experience is more important than distant grade-specific experience, I use variation in the timing of specific human capital acquisition. Because no baseline effects were found for reading, all analyses moving forward focus on math score improvement as the outcome.

I investigate decay using three distinct approaches. In the first specification, I interact grade-specific experience with whether the experience was accrued in the past five years. This specification tests whether distant grade-specific experience is important after controlling for recent grade-specific experience. The second approach explores the impact of having taught the same grade in the previous year, two years ago, three years ago, etc. In the third approach, I estimate a nonlinear model in which human capital is allowed to decay each period at a constant rate. The rate of decay is estimated along with the other parameters via NLS.

Column 1 of Table 5 shows results when the timing of grade-specific human capital is considered. All controls are identical to those in the main specification, but grade-specific experience is divided into recent (within five years) and distant (more than five years ago) experience. This specification suggests that recent grade-specific experience is much more important than distant grade-specific experience.[11] The recent grade-specific experience dummy coefficients are larger than those in the earlier specification, whereas the distant grade-specific experience dummies are mostly statistically insignificant. The coefficient on three years of distant grade-specific experience is negative and statistically significant, but this appears to be an anomalous result.

[11] These results are not sensitive to the exact definition of "recent."

TABLE 5—RECENT GRADE-SPECIFIC EXPERIENCE

	(1)	(2)		(3)	(4)
Accrued in past 5 years					
Expgrd=1	0.0143***	0.0243***	Same grade as $t-1$	0.0078	0.0205***
	(0.0054)	(0.0031)		(0.0124)	(0.0066)
Expgrd=2	0.0305***	0.0374***	Same grade as $t-2$		0.0158**
	(0.0066)	(0.0037)			(0.0065)
Expgrd=3	0.0375***	0.0392***	Same grade as $t-3$		0.0000
	(0.0079)	(0.0043)			(0.0065)
Expgrd=4	0.0471***	0.0467***	Same grade as $t-4$		0.0125**
	(0.0093)	(0.0048)			(0.0064)
Expgrd=5	0.0347***	0.0438***	Same grade as $t-5$		−0.0001
	(0.0106)	(0.0049)			(0.0056)
Accrued more than 5 years ago					
Expgrd=1	0.0023		*Expgrd*=1	0.0056	
	(0.0109)			(0.0131)	
Expgrd=2	−0.0219		*Expgrd*=2	0.0225	
	(0.0135)			(0.0139)	
Expgrd=3	−0.0373*		*Expgrd*=3	0.0270†	
	(0.0159)			(0.0148)	
Expgrd=4	−0.0051		*Expgrd*=4	0.0329*	
	(0.0169)			(0.0158)	
Expgrd=5	−0.0252		*Expgrd*=5	0.0242	
	(0.0199)			(0.0170)	
Expgrd=6	−0.0023		*Expgrd*=6	0.0311†	
	(0.0257)			(0.0184)	
Expgrd=7	−0.0062		*Expgrd*=7	0.0069	
	(0.0437)			(0.0203)	
Expgrd $\geq$ 8	−0.0102		*Expgrd* $\geq$ 8	0.0016	
	(0.0286)			(0.0211)	
Observations	507,017	1,170,686		365,181	484,856

Notes: The dependent variable is a standardized measure of test score. Each specification also controls for general experience dummies and all student and class controls from equation (1). Standard errors clustered at class level reported in parentheses.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Given that recent grade-specific experience appears to be most important, it is likely that little is lost by focusing on grade-specific experience acquired in the past five years. The advantage of doing so is that the main specification requires the entire history of teacher grade assignments so any teacher with out-of-sample experience must be dropped. Focusing on recently acquired grade-specific experience allows me to remove this sample restriction and more than double the sample size. Column 2 of Table 5 shows the result of estimating equation (1) where grade-specific experience is only counted if accrued in the past five years, and teachers who began prior to 1995 are included.[12] The magnitude of the dummy coefficients appears to

[12]In this specification, teacher-year observations are dropped if "five years prior" dates back before 1995 and the teacher was teaching prior to 1995. In general, each specification in this section uses the largest possible sample.

be slightly larger than those from the baseline specification, but generally, using the entire sample and considering only recent experience yields fairly similar results.

Column 3 adds an indicator for grade repetition to the baseline specification. One issue with investigating the importance of the preceding year's grade assignment is that teachers in their first year of teaching have no prior grade assignment, so these teachers cannot be included in the regression. As shown in column 3 of Table 5, controlling for grade repetition attenuates the estimated benefit of grade-specific experience somewhat, but the standard errors more than double, making several coefficients insignificant. The coefficient on the grade-repetition covariate itself is statistically insignificant and of small magnitude. When the grade-specific human capital controls are omitted, the grade-repetition coefficient increases and becomes statistically significant, suggesting that perhaps grade-specific experience and grade repetition are too colinear to disentangle their separate effects.

To provide further evidence regarding how the timing of grade-specific human capital matters, I replace the level of grade-specific experience with indicators for whether the same grade was taught one year ago, two years ago, etc. As in column 3, this specification drops inexperienced teacher-years since the key independent variables are undefined for these teachers; however, as in column 2, teachers who began prior to 1995 are included in the regression. Column 4 of Table 5 shows that repeating a grade assignment from the past two years benefits teacher performance, but repeating grade-assignments from earlier years has a diminishing additional benefit.

In the third and final approach to exploring depreciation, I explicitly model depreciation parametrically, and estimate the depreciation parameter. In the baseline specification, I had assumed that the level of grade-specific human capital is simply the number of years of grade-specific experience. To allow for depreciation, I weaken this assumption by specifying the evolution of human capital as:

$$K_{gt} = \mathbf{1}(g_t = g_{t-1}) + \beta K_{g,t-1}, \tag{2}$$

where $\mathbf{1}(g_t = g_{t-1})$ indicates that last year's grade assignment matches this year's grade assignment.[13] The parameter β is allowed to vary freely as opposed to constraining $\beta = 1$ as in the baseline specification.

Because the nonlinearity of the model makes estimation with half a million observations and thousands of fixed effects intractable, I reduce the dimensionality of the problem by estimating the model at the teacher-by-year level, rather than at the student level.[14] To remove unobserved teacher heterogeneity, I demean the teacher-by-year value-added by teacher prior to estimating the regression. While not literally equivalent to a teacher fixed effects specification, the interpretation of this model is very similar; namely, both relate changes in teacher performance to experience and grade-specific experience.

[13] Importantly, the g subscript refers to the grade taught in year t, not the grade taught in year $t-1$. As such, $K_{g,t-1}$ is not necessarily equal to last year's $K_{g,t}$, since the grade taught could be different in the two years.

[14] The teacher-by-year value-added measure is obtained by regressing student test scores on lagged test scores, student covariates, year-by-grade fixed effects and a teacher-by-year fixed effect. The results from Table 4 can be estimated at the teacher-by-year level as well and the results are very similar to estimating these models at the student level.

Using the demeaned teacher-by-year value-added measure as the dependent variable, I estimate:

$$VA_{jt} = f(Exp_{jt}) + h\left(\sum_{p=1}^{t} \mathbf{1}(g_t = g_{t-p})\beta^{p-1}\right) + \varepsilon_{ijgst}, \tag{3}$$

where $h(\cdot)$ is a function of grade-specific human capital and $f(\cdot)$ is a series of experience dummies. The quantity inside the summation is simply a discounted sum of all past years of grade-specific experience, such that if $\beta = 1$, the model is equivalent to that of equation (1). Since specific human capital is continuous in this context, $h(\cdot)$ cannot be specified as a series of indicators, and it is necessary to use a particular functional form. Since teacher experience effects have been shown to be highly nonlinear, I opt to use as flexible a functional form as possible. While I estimate the model using a linear, quadratic, cubic, quartic and quintic, I strongly prefer the more flexible parametric specifications since failure to capture the nonlinearity of the experience profile could lead to biased estimation of the depreciation parameter.

Row 1 of Table 6 shows the results of estimating equation (3) via NLS. The estimated depreciation rate is somewhat sensitive to the functional form of specific human capital, particularly when comparing between the linear and quadratic specifications. When using a cubic, quartic, or quintic specification, however, the estimated parameter is more consistent and implies a depreciation rate between 0.32 and 0.38 (only the cubic specification is shown in the table). Using even higher order polynomials does not change this basic estimate, suggesting that the depreciation rate estimates are not simply capturing nonlinearities.

Rows 2 through 5 of Table 6 show predicted productivity levels of teachers with various levels of depreciated experience. These predicted values are presented to ease the interpretation of the complex higher order polynomials. While the exact magnitudes vary somewhat depending on the polynomial order, teachers with five years of grade-specific experience are predicted to perform approximately 0.026 standard deviations better than a novice teacher, but this estimate falls substantially if the grade-specific experience has decayed for two years. In the baseline specification, the estimated benefit of five years of grade-specific experience was 0.032, which is larger than the parametric estimates, but the same order of magnitude.

Estimates of human capital depreciation from the labor literature are generally in the 0.01 to 0.05 range (Mincer and Ofek 1982; Görlich and de Grip 2009), and it is worth discussing whether it is plausible that the teacher productivity-based measure would yield such a different estimate. First, it is worth noting that studies of teacher improvement have generally found results that differ dramatically from wage-based studies. For example, a consistent finding across many studies is that teachers make more improvement in their first year than in all future years combined (Clotfelter, Ladd, and Vigdor 2007; Hanushek et al. 2005; Rivkin, Hanushek, and Kain 2005). Wage experience profiles, on the other hand, increase at a much more gradual and sustained rate. The divergence between the wage-based estimate and the teacher productivity based estimates could be because teachers improve very differently than other workers, or it could be that wages in general do not perfectly reflect workers' instantaneous productivity. For example, a model of wage-deferring contracts

Table 6—Human Capital Depreciation and Heterogeneity

Functional form		Linear	Quadratic	Cubic
Panel A. Parametric depreciation estimates				
Estimate of β		0.3820**	0.8025***	0.6814***
		(0.1697)	(0.0909)	(0.0979)
Implied benefit of 2 years grd-spec exp.				
Accrued in past 2 years		0.0189	0.0213	0.0207
Accrued with 2 year delay		0.0028	0.0154	0.0059
Implied benefit of 5 years grd-spec. exp.				
Accrued in past 5 years		0.0240	0.0270	0.0258
Accrued with 2 year delay		0.0035	0.0265	0.0165
		By years of experience		
Sample	All experience levels (1)	*Exp.* $\leq$ 10 (2)	10 $<$ *Exp.* $\leq$ 20 (3)	*Exp.* $>$ 20 (4)
Panel B. Heterogeneity by experience				
Expgrd=1	0.0243***	0.0233***	0.0170**	0.0179*
	(0.0031)	(0.0037)	(0.0081)	(0.0103)
Expgrd=2	0.0374***	0.0398***	0.0295***	0.0080
	(0.0037)	(0.0047)	(0.0093)	(0.0104)
Expgrd=3	0.0392***	0.0424***	0.0272***	0.0119
	(0.0043)	(0.0059)	(0.0095)	(0.0111)
Expgrd=4	0.0467***	0.0510***	0.0374***	0.0158
	(0.0048)	(.0073)	(0.0103)	(0.0110)
Expgrd=5	0.0438***	0.0366***	0.0354***	0.0274***
	(0.0049)	(0.0088)	(0.0104)	(0.0105)
Observations	1,170,686	701,076	208,834	260,776

Notes: The dependent variable is a standardized measure of test score. Each specification also controls for general experience dummies and all student and class controls from equation (1).

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

would predict a smooth experience profile, even if productivity is more or less variable (Lazear 1979). Second, while 0.35 may seem like a fast rate of decay, it is not implausibly large when considered in conjunction with the nonparametric specification shown in column 1 of Table 5. That specification shows that grade-specific experience acquired more than five years ago yields no added benefit, and were the depreciation rate in the 0.01–0.05 range, experience from five years ago would still be expected to yield benefits.

B. *Heterogeneity by Experience Level*

Theoretically, grade-specific experience could be more or less useful depending on one's level of general human capital. If the two inputs are complements, then teachers with many years of general experience would be able to better use their grade-specific human capital, whereas if the two inputs are substitutes, grade-specific human capital would be most useful for inexperienced teachers. Panel B of Table 6 explores whether the rate of grade-specific improvement varies with one's level of general experience. In order to be able to include very experienced teachers, this specification focuses on grade-specific human capital accrued in the past five years and teachers are included even if they began teaching prior to 1995.

Comparing columns 2 and 3 of panel B shows that both relatively inexperienced and moderately experienced teachers benefit from grade-specific experience, but the benefits are slightly larger for less experienced teachers. Column 4 of panel B shows that for very experienced teachers, the benefits of grade-specific experience are substantially smaller. With the exception of the first year of grade-specific experience, the magnitude of the effects are smaller for very experienced teachers and the overall pattern is much weaker. Note that the lack of statistical significance for the very experienced sample is not driven by large standard errors since the standard errors in columns 3 and 4 are of comparable magnitude. In results not shown, I find that very experienced teachers similarly gain little to no benefit from having taught their current grade in the recent past. These results suggest that grade-specific human capital and general teaching human capital are substitutes.

VI. Identification Tests

When including teacher fixed effects, the impact of specific experience is identified by the divergence between grade-specific experience and general experience that occurs due to grade switching. Estimates will be biased if teachers are switched in a way that is systematically related to their expected performance.[15] While the variation that identifies the return to grade-specific experience is generated by grade-switching, occasionally grade switches result in an increase in grade-specific experience. As such, I test both whether dynamic performance predicts grade switching, and also whether dynamic performance predicts changes in grade-specific experience across years.

A. *Test for Endogenous Grade Switching*

Assuming some mean reversion, a teacher who is switched after a particularly poor year will tend to do better in the year after the switch compared to the year before the switch. If teachers are systematically switched after particularly bad years, this will result in understating the benefits of grade-specific experience, whereas if teachers are switched after particularly good years, this will upwardly bias the estimated returns to grade-specific experience. As such, the first proxy for expected performance in year $t + 1$ is to test whether performance in year t is systematically related to switching grades between year t and year $t + 1$.

I test for switching based on dynamic performance by estimating equation (4), which predicts whether a teacher is switched between year t and $t + 1$.

$$\mathbf{1}(g_{j,t} \neq g_{j,t+1}) = \gamma \overline{\Delta A}_{j_t g_t s_t t} + \beta \overline{\mathbf{X}}_{jt} + \sigma \mathbf{R}_{gst} + \xi_{gt} + \omega_j + \varepsilon_{jt}. \tag{4}$$

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

The variable $\mathbf{1}(g_{j,t} \neq g_{j,t+1})$ is an indicator that is unity when teacher j switches grade assignments and zero when teacher j repeats grade assignments. To test whether teachers are switched due to recent performance, I include a control for the gains made by students in the teacher's time t classroom. If teachers are switched after years in which their students perform particularly badly, the coefficient γ will be negative. The vector $\overline{\mathbf{X}}_{jt}$ includes the average student characteristics taught by teacher j in year t. For ease of interpretation, all class-level covariates are standardized by dividing by their standard deviation, thus the estimated coefficients correspond to a one standard deviation change. The vector $\mathbf{R}_{gst}$ indicates whether the number of sections of grade g increased or decreased between year t and year $t+1$.[16] All other variables in equation (4) are defined as in equation (1). This model is run as a linear probability model rather than a nonlinear model for simplicity and because empirically, predicted probabilities all lie between zero and one using the LPM.

Column 1 of Table 7 shows the key results, controlling for teacher fixed effects. First, it is apparent that section offering changes have an asymmetric impact on teacher switching. When the number of sections is reduced, this increases the likelihood that a teacher is switched by 6.39 percentage points—around a 30 percent increase. When the number of sections is increased, however, this has no impact on the likelihood that a teacher is switched. This asymmetry is to be expected since when the number of sections offered for a grade is reduced, unless a teacher quits, one of the teachers who previously taught the grade will need to be switched. When the number of sections offered is increased, on the other hand, this could plausibly make it less likely that a teacher will switch away from this grade, but the mechanism is less direct.

To test for dynamic endogenous switching, the key covariate of interest is whether teachers switch following years in which they perform particularly well or badly. Column 1 of Table 7 shows that there is no evidence that teachers are switched based on past performance.

Because of the nature of the specification test, teachers must be observed in two consecutive years to be included in the regression. This restriction leads to a considerable sample size reduction, which lowers the power of these regressions. Regardless, in terms of magnitude, the point estimates are quite small and their signs vary across math and reading scores. A one standard deviation increase in a student's math test score corresponds to no change in the probability that a teacher will switch grades, while a one standard deviation increase in reading score gains corresponds to a 0.02 percent increase in the probability that a teacher will switch grades. While dynamic endogenous switching will only be a problem to the extent that it is correlated with teacher performance, column 1 of Table 7 also shows that there is little evidence of a relationship between teacher switching and the characteristics of students just taught.

[16] Hoxby (2000) exploits population variation to identify the effect of class size on student achievement. Similarly, population variation leads to changes in the number of sections 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.

TABLE 7—TEST FOR DYNAMIC ENDOGENEITY OF SWITCHING AND GRADE-SPECIFIC EXPERIENCE

Dependent variable	Probability teacher is switched between year t and $t+1$	Grade-specific experience in year $t+1$
Number of sections offered decreases between t and $t+1$	0.0639***	−0.1402***
	(0.00941)	(0.03495)
Number of sections offered increases between t and $t+1$	0.0014	−0.0216
	(0.00917)	(0.03454)
Standardized math gains in year t	−0.0000	0.0003*
	(0.00004)	(0.00015)
Standardized reading gains in year t	0.0002	−0.0009
	(0.00019)	(0.00073)
Standardized frac. female in year t	−0.0265	0.0834
	(0.01950)	(0.07230)
Standardized frac. black in year t	−0.0097	0.0137
	(0.01388)	(0.04998)
Standardized frac. Hispanic in year t	0.0052	0.0002
	(0.00563)	(0.02165)
Standardized frac. limited English in year t	0.0038	−0.0210
	(0.00378)	(0.01586)
Standardized frac. free lunch in year t	−0.0014	−0.0270
	(0.02000)	(0.07581)
Observations	17,547	17,547

Notes: For the first column, the dependent variable is an indicator for whether a teacher switches grades between year t and year $t+1$. In the second column, the dependent variable is the number of years of grade-specific experience a teacher has in year $t+1$. The first two covariates are indicators for whether the number of sections offered increases or decreases between year t and $t+1$. These variables are calculated at the school-year-grade level to account for the fact that a decrease in the number of sections of third grade increases the likelihood that third grade teachers will be switched, but does not impact a fourth grade teacher directly. The next seven covariates are average class characteristics in year t. For ease of interpretation, each of these variables is standardized by dividing by its standard deviation. Both specifications control for grade-by-year fixed effects and teacher fixed effects. The specification which predicts grade-specific experience in year $t+1$ conditions on grade-specific experience in t, so the regression captures whether each covariate impacts changes in grade-specific experience. Standard errors reported in parentheses.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

B. *Test for Endogenous Changes in Grade-Specific Experience*

If teacher expected performance is related to changes in grade-specific experience, then estimates of the returns to grade-specific experience will be biased. To test for this potential bias, I reestimate equation (4) but instead of predicting grade switches, the specification predicts years of grade-specific experience in year $t+1$ conditional on the amount of grade-specific experience in year t. Specifically, I estimate

$$(5) \qquad Expgrd_{jt+1} = g(Expgrd_{jt}) + \gamma \overline{\Delta A}_{j_t g_t s_t} + \beta \overline{\mathbf{X}}_{jt} + \sigma \mathbf{R}_{gst} + \xi_{gt} + \omega_j + \varepsilon_{jt},$$

where all covariates are defined as in equation (4).

Column 2 of Table 7 shows that the qualitative findings regarding grade switching carry over to predicting changes in grade-specific experience. There is one marginally statistically significant coefficient, but the magnitude of the effect is very small. A one standard deviation increase in average math score gains corresponds to a 0.0003 change in years of grade-specific experience.

Taken together, the results of Table 7 combined with the results shown in Table 4 suggest that particular types of teachers may be switched more often than others, but the timing of when a teacher is switched is not correlated with dynamic aspects of her performance.

C. *Test for Student Sorting*

The tests shown in the previous section aimed to distinguish whether teachers improve with grade-specific experience or simply that better performing teachers also happen to have more grade-specific experience. Another possibility is that teachers are simply given better students as they gain grade-specific experience, and what appears to be teacher improvement could just be student sorting. To examine whether the experience improvement profile shown in Table 4 is driven by more experienced teachers getting better students, I show how a teacher's experience level predicts student lagged test scores. This test is similar to that proposed by Rothstein (2010).

Table 8 shows how a teacher's experience level corresponds to both the incoming and outgoing test scores of her students. These specifications are identical to those from the main specification, but column 1 predicts lagged test scores instead of current test scores, and column 2 predicts current scores but omits the lagged test score control. Column 1 shows that teachers with more experience are given students with higher past test scores, but teachers are not given different students according to their level of grade-specific experience. To ascertain whether the improvement profile can be explained away by student sorting, I compare the results for the incoming and outgoing test scores. The incoming test scores cannot be impacted by the teacher and thus significant effects in column 1 must reflect student sorting. The outgoing test scores on the other hand are reflective of both the types of students assigned to teacher j and also what they have learned from teacher j. If the incoming and outgoing scores are of a similar magnitude, then this would suggest that student sorting accounts for the experience profile and teachers do not improve with experience.

Comparing columns 1 and 2 of Table 8 demonstrates two facts. First, there is no difference in incoming test scores according to a teacher's grade-specific experience, but the outgoing test scores are higher for students taught by teachers with more grade-specific experience. Second, teachers are assigned better students as they gain experience, but this sorting does not explain away the improvement effects. For example, the incoming test scores of students assigned to a second year teacher are 0.0326 standard deviations better than a novice teachers' students, but the outgoing test scores are 0.0910 standard deviations better. This suggests that the more experienced teachers get somewhat better students, but it remains the case that students benefit from a teacher's experience level.

Table 8—Student Sorting to Teachers

Model	Predicting test score in year $t-1$ (1)	Predicting test score in year t (2)	Variation at school-by-grade-by-year level (3)	School-grade-year IV (4)
Expgrd=1	0.0013	0.0154**	0.0256***	0.0136***
	(0.0063)	(0.0070)	(0.0065)	(0.0043)
Expgrd=2	−0.0020	0.0280***	0.0407***	0.0276***
	(0.0076)	(0.0086)	(0.0074)	(0.0051)
Expgrd=3	−0.0112	0.0229**	0.0687***	0.0321***
	(0.0093)	(0.0105)	(0.0083)	(0.0061)
Expgrd=4	−0.0051	0.0373***	0.0824***	0.0408***
	(0.0109)	(0.0124)	(0.0097)	(0.0071)
Expgrd=5	−0.0026	0.0316**	0.0560***	0.0287***
	(0.0131)	(0.0146)	(0.0112)	(0.0082)
Expgrd=6	0.0122	0.0494***	0.1037***	0.0428***
	(0.0154)	(0.0173)	(0.0133)	(0.0096)
Expgrd=7	0.0212	0.0365*	0.1167***	0.0269**
	(0.0179)	(0.0202)	(0.0153)	(0.0112)
Expgrd $\geq$ 8	−0.0208	−0.0094	0.1152***	0.0035
	(0.0186)	(0.0214)	(0.0137)	(0.0116)
Exp=1	0.0106	0.0599***	0.0564***	0.0497***
	(0.0069)	(0.0078)	(0.0071)	(0.0048)
Exp=2	0.0326***	0.0910***	0.0771***	0.0649***
	(0.0081)	(0.0092)	(0.0078)	(0.0056)
Exp=3	0.0416***	0.1068***	0.0750***	0.0666***
	(0.0096)	(0.0108)	(0.0083)	(0.0064)
Exp=4	0.0463***	0.1046***	0.0647***	0.0590***
	(0.0109)	(0.0123)	(0.0091)	(0.0071)
Exp=5	0.0498***	0.1156***	0.0762***	0.0716***
	(0.0126)	(0.0140)	(0.0101)	(0.0080)
Exp=6	0.0515***	0.1197***	0.0675***	0.0637***
	(0.0141)	(0.0157)	(0.0115)	(0.0089)
Exp=7	0.0430***	0.1136***	0.0811***	0.0636***
	(0.0157)	(0.0180)	(0.0129)	(0.0101)
Exp $\geq$ 8	0.0546***	0.1463***	0.0872***	0.0812***
	(0.0162)	(0.0183)	(0.0108)	(0.0104)
Fixed effect	Teacher	Teacher	School-year	Teacher
Observations	507,746	507,011	510,216	507,193

Notes: The dependent variable in column 1 is lagged test score. Column 2 predicts test score at the end of year *t*, without conditioning on lagged test score. Columns 3 and 4 use school-by-grade-by-year variation to circumvent the potential problem of student sorting. For column 3, the independent variables correspond to the proportion of teachers in a particular school-grade-year who have a given level of experience. All specifications include all of the same controls as in equation (1). See text for more details. Standard errors clustered at class level reported in parentheses.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

To further explore whether the general experience profile is driven by student sorting, I follow Rivkin, Hanushek, and Kain (2005) and aggregate the level of analysis to the school-grade-year. The idea behind this aggregation is that it removes any student sorting that occurs within a school-grade-year. If the most promising students in a particular grade are sorted towards the most experienced teachers, this would bias estimates of the experience effects, but aggregating to the school-grade-year level removes this form of sorting. When aggregated to the school-grade-year level, rather than examining the impact of having a teacher with more years of experience, I examine the impact of being in a grade with a higher fraction of more experienced teachers. To maintain the flexibility of earlier specifications, I include separate variables indicating the fraction of teachers in the grade with one year of experience, two years of experience, etc.

Column 3 of Table 8 shows results when the analysis is aggregated to the school-year-grade level. These results clearly demonstrate that students in grades that have more experienced teachers make larger gains than grades with less experienced teachers. While this suggests that student sorting did not bias previous results, one drawback with the Rivkin, Hanushek, and Kain (2005) methodology is that it is not possible to directly include teacher fixed effects, which were previously shown to be important for evaluating the benefits of experience. In order to be able to include teacher fixed effects, I slightly extend the Rivkin, Hanushek, and Kain (2005) methodology by using school-grade-year variation as an instrument for teacher experience. By placing the analysis in the IV context, I am able to use school-grade-year level variation for identification, while still including teacher fixed effects. I use 16 separate instruments, each one corresponding to a particular experience or grade-specific experience level. For example, to instrument for whether a teacher has threes years of experience, I use the fraction of teachers in her school-grade-year who have three years of experience. Column 4 of Table 8 shows the results when a teacher's level of experience is instrumented for using the fraction of teachers in her school-grade-year with that particular level of experience. Compared to column 3, the coefficients in column 4 are considerably smaller, which highlights the importance of controlling for the teacher fixed effects. With the teacher fixed effect controlled for, column 4 shows that using school-grade-year level variation yields very similar estimates to those from the main specification (column 2 of Table 4).

Based on these identification tests, I conclude that specifications that include teacher fixed effects likely provide reasonable estimates of the impact of grade-specific experience.

VII. Conclusion

While panel data has been used extensively to make methodological improvements to research on teacher quality, relatively few studies have used these data to analyze in detail how the past impacts the future. This study measures not only whether a teacher was teaching five years earlier, but considers *what* she was teaching. Using this dynamic measure of task assignments within a school, this paper separately identifies the productivity benefits of general and grade-specific human capital and finds that both are important in determining the rate of teacher improvement.

While I find that grade-specific human capital is beneficial, these benefits depreciate rapidly. As a result, switching teacher grade assignments has modest short-term negative effects, but few long-term consequences for a teacher's productivity. That said, the extent of grade switching documented in this paper suggests that students are frequently exposed to a teacher who just switched grade assignments. Furthermore, since grade-specific human capital is found to be most important for inexperienced teachers, students at schools staffed by inexperienced teachers are disproportionately impacted by grade-switching.

The fact that teachers improve with experience is commonly cited as one reason why teacher attrition is problematic. This paper shows that frequently 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. 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. Furthermore, since I find that the value-added of very experienced teachers is unaffected by grade switches, it might be best to switch more experienced teachers, all else equal.

Appendix

Table A1—Robustness of Main Result to Type of Value-Added Model

	Math			Reading		
Model	Preferred specification (1)	Gains model (2)	Lagged IV (3)	Preferred specification (4)	Gains model (5)	Lagged IV (6)
Expgrd=1	0.0136*	0.0103†	0.0162†	−0.0045	0.0064	−0.0035
	(0.0054)	(0.0055)	(0.0083)	(0.0050)	(0.0056)	(0.0095)
Expgrd=2	0.0292***	0.0257***	0.0363***	−0.0005	0.0168**	0.0164
	(0.0066)	(0.0067)	(0.0112)	(0.0061)	(0.0068)	(0.0126)
Expgrd=3	0.0361***	0.0325***	0.0327**	0.0099	0.0189**	0.0201
	(0.0080)	(0.0081)	(0.0142)	(0.0073)	(0.0082)	(0.0160)
Expgrd=4	0.0422***	0.0351***	0.0432**	0.0059	0.0225**	−0.0076
	(0.0093)	(0.0095)	(0.0170)	(0.0085)	(0.0096)	(0.0193)
Expgrd=5	0.0357***	0.0309***	0.0328	0.0021	0.0195*	0.0109
	(0.0111)	(0.0113)	(0.0201)	(0.0101)	(0.0114)	(0.0228)
Expgrd=6	0.0412***	0.0377***	0.0342	−0.0015	0.0270**	0.0187
	(0.0129)	(0.0132)	(0.0233)	(0.0115)	(0.0134)	(0.0267)
Expgrd=7	0.0209	0.0165	0.0199	−0.0120	0.0089	0.0052
	(0.0154)	(0.0157)	(0.0270)	(0.0137)	(0.0156)	(0.0307)
Expgrd ≥ 8	0.0152	0.0172	0.0150	−0.0141	0.0044	0.0261
	(0.0160)	(0.0164)	(0.0314)	(0.0137)	(0.0160)	(0.0349)
Exp=1	0.0468***	0.0553***	0.0344***	0.0320***	0.0520***	0.0316***
	(0.0060)	(0.0068)	(0.0095)	(0.0054)	(0.0069)	(0.0109)
Exp=2	0.0555***	0.0683***	0.0259**	0.0489***	0.0612***	0.0318**
	(0.0071)	(0.0093)	(0.0125)	(0.0065)	(0.0094)	(0.0142)
Exp=3	0.0581***	0.0776***	0.0374**	0.0393***	0.0661***	0.0332*
	(0.0083)	(0.0122)	(0.0158)	(0.0075)	(0.0123)	(0.0181)
Exp=4	0.0507***	0.0780***	0.0257	0.0490***	0.0619***	0.0572***
	(0.0093)	(0.0151)	(0.0189)	(0.0085)	(0.0152)	(0.0217)
Exp=5	0.0563***	0.0863***	0.0272	0.0453***	0.0644***	0.0233
	(0.0107)	(0.0183)	(0.0223)	(0.0095)	(0.0184)	(0.0258)
Exp=6	0.0562***	0.0944***	0.0288	0.0412***	0.0628***	0.0248
	(0.0116)	(0.0216)	(0.0259)	(0.0104)	(0.0215)	(0.0300)
Exp=7	0.0552***	0.0902***	0.0059	0.0497***	0.0626**	0.0436
	(0.0135)	(0.0250)	(0.0295)	(0.0124)	(0.0253)	(0.0344)
Exp ≥ 8	0.0695***	0.1207***	0.0450	0.0596***	0.0794***	0.0292
	(0.0140)	(0.0307)	(0.0351)	(0.0122)	(0.0306)	(0.0406)
Observations	507,017	507,017	168,152	497,888	498,729	164,126

Notes: The dependent variable is a standardized measure of test score. For comparison purposes, columns 1 and 4 simply duplicate the main results from Table 4. Columns 2 and 5 show analogous results using a simple gains model. Columns 3 and 6 use the second lag to instrument for the first lag and consequently drop any student without three consecutive years of test score data. Standard errors clustered at class level reported in parentheses.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

REFERENCES

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. “Teachers and Student Achievement in the Chicago Public High Schools.” *Journal of Labor Economics* 25 (1): 95–135.

Anderson, T. W., and Cheng Hsiao. 1981. “Estimation of dynamic models with error components.” *Journal of the American Statistical Association* 76 (375): 598–606.

Becker, Gary S. 1964. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. Chicago: University of Chicago Press.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2011. “The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood.” National Bureau of Economic Research (NBER) Working Paper 17699.

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): 673–82.

Dee, Thomas S. 2005. “A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?” *American Economic Review* 95 (2): 158–65.

Görlich, Dennis, and Andries de Grip. 2009. “Human capital depreciation during hometime.” *Oxford Economic Papers* 61 (1): 98–121.

Hanushek, Eric A., John F. Kain, Daniel M. O’Brien, and Steven G. Rivkin. 2005. “The Market for Teacher Quality.” National Bureau of Economic Research (NBER) Working Paper 11154.

Harris, Douglas N. 2009. “Would Accountability Based on Teacher Value Added Be Smart Policy? An Examination of the Statistical Properties and Policy Alternatives.” *Education Finance and Policy* 4 (4): 319–50.

Hoxby, Caroline M. 2000. “The Effects of Class Size on Student Achievement: New Evidence from Population Variation.” *Quarterly Journal of Economics* 115 (4): 1239–85.

Hoxby, Caroline M., and Sonali Murarka. 2009. “Charter Schools in New York City: Who Enrolls and How They Affect Their Students’ Achievement.” National Bureau of Economic Research (NBER) Working Paper 14852.

Jackson, C. 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., Stephanie K. Riegg, and Douglas O. Staiger. 2006. “School quality, neighborhoods, and housing prices.” *American Law and Economics Review* 8 (2): 183–212.

Kane, Thomas J., and Douglas O. Staiger. 2008. “Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation.” National Bureau of Economic Research (NBER) Working Paper 14607.

Koedel, Cory, and Julian Betts. 2010. “Value Added to What? How a Ceiling in the Testing Instrument Influences Value-Added Estimation.” *Education Finance and Policy* 5 (1): 54–81.

Lazear, Edward P. 1979. “Why Is There Mandatory Retirement?” *Journal of Political Economy* 87 (6): 1261–84.

Mincer, Jacob, and Haim Ofek. 1982. “Interrupted Work Careers: Depreciation and Restoration of Human Capital.” *Journal of Human Resources* 17 (1): 3–24.

North Carolina Department of Education. 2009. “Accountability Services.” http://www.dpi.state.nc.us/accountability/testing/eog/ (accessed May 1, 2009).

North Carolina Education Research Data Center. 1995–2012. “Student, Class and Personnel Files.” http://childandfamilypolicy.duke.edu/research/hc-education-data-center/.

Ost, Ben. 2014. “How Do Teachers Improve? The Relative Importance of Specific and General Human Capital: Dataset.” *American Economic Journal: Applied Economics.* http://dx.doi.org/10.1257/app.6.2.127.

Papay, John P., and Matthew Kraft. 2013. “Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-Term Career Improvement.” Working Paper.

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

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

Rothstein, Jesse. 2010. “Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement.” *Quarterly Journal of Economics* 125 (1): 175–214.

Todd, Petra E., and Kenneth I. Wolpin. 2003. “On the Specification and Estimation of the Production Function for Cognitive Achievement.” *Economic Journal* 113 (485): F3–33.