RETURNS TO TEACHER EXPERIENCE: STUDENT ACHIEVEMENT AND MOTIVATION IN MIDDLE SCHOOL

Helen F. Ladd
Sanford School of Public Policy
Duke University and CALDER
Durham, NC 27708-0243
hladd@duke.edu

Lucy C. Sorensen
(corresponding author)
Rockefeller College of Public Affairs and Policy
University of Albany SUNY
Albany, NY 12222
lsorensen@albany.edu

Abstract
We use rich longitudinally matched administrative data on students and teachers in North Carolina to examine the patterns of differential effectiveness by teachers' years of experience. The paper contributes to the literature by focusing on middle school teachers and by extending the analysis to student outcomes beyond test scores. Once we control statistically for the quality of individual teachers using teacher fixed effects, we find large returns to experience for middle school teachers in the form both of higher test scores and improvements in student behavior, with the clearest behavioral effects emerging for reductions in student absenteeism. Moreover these returns extend well beyond the first few years of teaching. The paper contributes to policy debates by documenting that teachers can and do continue to learn on the job.

doi:10.1162/EDFP_a_00194

1. INTRODUCTION

In the late 1980s, most of the nation's teachers had considerable experience—only 17 percent had taught for five or fewer years. By 2008, however, about 28 percent—or more than one in four of America's teachers—had five or fewer years of experience (Ingersoll and Merrill 2010). Some observers applaud this "greening" of the teaching force because they believe that experience is not crucial to teacher effectiveness. Others decry the trend because they believe that good teaching is a complex process that can best be mastered through years of classroom experience. The purpose of this paper is to speak to this policy debate. In particular, we use longitudinally matched administrative data on middle school students and teachers in North Carolina to examine the patterns of differential effectiveness by teachers' years of experience.

Most researchers and policy makers agree that regardless of how effective they may eventually become, novice teachers are typically less effective than their counterparts who have more experience. Under debate is the extent to which teachers continue to learn on the job beyond the first few years of teaching. In recent years, researchers have used administrative data from individual states and districts to examine how teaching experience contributes to student test scores. Research by Kane, Rockoff, and Staiger (2008) concludes, for example, that although teachers have a steep learning curve for the first few years, the test score returns to experience level off after four or five years. Using data from New York City, Boyd et al. (2008) report a similar pattern for both elementary and middle school math teachers. Although studies by Clotfelter, Ladd, and Vigdor, based on North Carolina data (2006, 2007a, and 2010), confirm that the returns to experience are largest in the first five years, they report some limited gains beyond that for both elementary school math and reading teachers, but none for high school teachers. Using Florida data, Harris and Sass (2011) find rising returns to experience through more than twenty years of experience for elementary and secondary school teachers, but negative returns for high school teachers. The study by Aaronson, Barrow, and Sander (2007), which focuses on ninth grade teachers in Chicago, is one of the few studies to report no returns to experience.

The patterns undoubtedly differ across these and other studies in part because of the differing contexts. Another reason they may differ, however, is that researchers have used different statistical methods, ranging from cross-sectional analysis (Clotfelter, Ladd, and Vigdor 2006) to panel data with various fixed effects (Hanushek et al. 2005; Clotfelter, Ladd, and Vigdor 2007a), to two-stage models (Aaronson, Barrow, and Sander 2007), to random assignment and natural experiments (Dee 2004). In addition, researchers specify the experience variable in different ways. Some researchers group experience levels into categories (Clotfelter, Ladd, and Vigdor 2006, 2007a, 2010), some constrain returns at higher levels to be zero (Rockoff 2004), and some use a detailed set of indicators (Boyd et al. 2008). Two recent papers (Wiswall 2013; Papay and Kraft 2015) have drawn close attention to the specification of the experience variables and to the identifying assumptions that underlie them. As we explain in the following, we combine the insights from these two recent papers with other modeling considerations for this paper. Hence, the first contribution of this paper is our careful attention to model specification.

A second contribution is our focus on returns to teacher experience in middle schools, an understudied level of schooling. In their 2011 summary of research on teacher experience, Harris and Sass (2011) include 22 estimates of experience effects for elementary school teachers (12 based on math test scores and 10 on reading scores) but only 7 for middle school teachers (4 for math and 3 for reading) and only 4 for high school teachers (2 for math, 1 for reading, and 1 for combined math and reading scores). Although the federal No Child Left Behind Act required every state to test all students in math and reading annually in grades 3–8, several studies focus on teachers in the elementary grades alone. That is especially true for the many studies that rely on North Carolina statewide data because, until recently, it was not possible to match students to their teachers in middle schools in that state (Clotfelter, Ladd, and Vigdor 2007b; Harris and Sass 2011; Wiswall 2013; Xu, Ozek, and Hansen 2014).

Starting in 2006, however, the North Carolina Department of Public Instruction has provided administrative data to the North Carolina Education Research Data Center in a form that links students to their teachers. As a result, we are able to use panel data from 2007 to 2011 to examine North Carolina middle school teachers. For the purposes of this study, we define middle school students as those in grades 6, 7, or 8 regardless of the type of school (e.g., middle, elementary, or K–12) that offers the grade. Given the differing existing findings related to experience for teachers at the elementary and high school levels, and the special position of middle schools between the nurturing environment of elementary schools and the larger size and departmentalization of high schools, more attention to teachers at the middle school level is desirable.

A third contribution of the paper is that we broaden the types of returns to teacher experience beyond student test scores to include student behaviors. Clearly, effective teachers do more for students on a daily basis than simply imparting a narrow set of reading or math skills. Ideally, such teachers cultivate character, discipline, and curiosity, and a variety of other capacities that are sometimes referred to as noncognitive skills. Research has documented that noncognitive capabilities in adolescence strongly influence educational attainment, employment, earnings, occupation, antisocial behavior, and substance use in adulthood (Heckman, Stixrud, and Urzua 2006; Lleras 2008). We build here on the work of Jackson (2012a), who uses data on ninth grade teachers in North Carolina, combined with longitudinal survey data on the relationship between noncognitive skills and long-run outcomes, to demonstrate that teachers contribute to the long-run well-being of students not only by raising student test scores but also by developing their noncognitive skills. Moreover, he shows that teachers who are productive in one dimension are not always productive in the other dimension.

As we will describe in more detail, we are able to estimate how a teacher's experience contributes to four non-test score outcome measures: absences, reported disruptive classroom offenses, time spent completing homework, and time spent reading for pleasure. Although these behaviors depend in part on a student's home or community environment, they also reflect important facets of learned motivation, perseverance, and self-control, and may be particularly important for the future success of middle school students. Evidence that more experienced teachers generate more positive outcomes along these dimensions as well as for test scores would greatly strengthen the policy argument for pursuing policies designed to develop and retain experienced teachers.

Based on our preferred models, we find clear returns to experience in the form of higher student test scores. These returns rise at least through twelve years of experience both for middle school teachers of math and English language arts (ELA). Consistent with prior research, the returns to experience are largest during the first few years, especially for math teachers. Contrary to the received wisdom, however, the returns continue to rise well beyond the first five years for teachers of both subjects. Although the returns level off after 12 years for math teachers, our results imply that math teachers with 21–27 years were still 0.04 standard deviations more effective than they would have been after five years. For ELA, teachers with 21–27 years of experience were also 0.04 standard deviations more effective as when they had five years of experience.

Despite somewhat more mixed patterns for the non-test outcomes (and less precision), some evidence of positive returns to experience still emerge. The most consistent findings appear for teachers' success in preventing student absences. In particular, experienced teachers effectively reduce the number of students with high levels of absenteeism across both ELA and math classrooms. In addition, experience appears to increase the ability of ELA teachers to encourage free reading and possibly time spent on homework (as reported by students). Also, weak evidence supports the view that math teachers become more effective in reducing disciplinary offenses, especially in the first several years of teaching. The statistical models on which these findings are based are similar to those used for the test scores, and hence address the various statistical challenges that arise in estimating any type of return to experience. We conclude that as individual teachers gain experience they become more effective not only in raising test scores but also in contributing to other valued behaviors, such as attending school or reading outside of school.

Of course not all teachers are equally effective—nor would we expect the returns to experience to be the same for all teachers in all school environments. Hence, the first task for policy makers is to recruit high-quality teachers. In light of our findings, the challenge then is to support their development and to retain them.

2. WHY PAY ATTENTION TO TEACHER EXPERIENCE IN MIDDLE SCHOOL?

Despite how simple good teaching may appear to the outside observer, extensive educational research, most of which is qualitative, shows that good teaching is in fact a complex process that poses challenges not present in other professions (Labaree 2000). The factors that make teaching difficult start with the problem of client cooperation. That is, the success of the teacher depends heavily on the active cooperation of the student. Cohen's description is appropriate when he describes students as people with their own motivations and wills and not inanimate objects like blocks of wood that are able to be sculpted by a carpenter (Cohen 1996). An associated factor is that the students are required to be in school, and many might not be there if they had a choice. Teachers also face the challenge of managing emotions. Further is the problem of structural isolation, as most teachers teach in self-contained classrooms, where the first order of business may be to maintain control. Finally, teachers have to deal with the chronic uncertainty about their effectiveness as a consequence of the challenges they face and the often contradictory purposes that societies impose on the educational endeavor.

Like experts in other domains, experienced teachers quickly recognize patterns in what they observe, see more complexities than novices, and bring to bear many sources of knowledge about how to respond to them. In addition, they are flexible in their practice and have a broad repertoire of skills with which they can easily access and implement to achieve their goals (Berliner 2001, 2004). The implication is that teachers can learn the basic tools of teaching in their pre-service training but mastery of teaching requires extensive reflection that can only come with experience and exposure to a variety of classroom experiences.

The process of teaching is likely to be particularly complex for middle school teachers because it happens amid a critical period of cognitive, socioemotional, and biological development of students who confront heightened social pressures from peers and gradual decline of parental oversight. As Laurence Steinberg summarizes, "adolescence is characterized by an increased need to regulate affect and behavior in accordance with long-term goals and consequences.... Because developing brain, behavioral and cognitive systems mature at different rates and under the control of both common and independent biological processes, this period is often one of increased vulnerability and adjustment" (Steinberg 2005, p. 69). The experience of transitioning to middle school can overwhelm students who often face a larger, more heterogeneous set of peers, lower levels of student engagement, and classroom teachers with less experience and higher turnover (Byrnes and Ruby 2007). Behaviors that emerge for some students in middle school—including chronic absenteeism, failed courses, and suspensions—provide a fairly reliable warning sign of dropping out of high school (Balfanz, Herzog, and Mac Iver 2007). If indeed middle school represents such a critical turning point for many students, positive teacher influence or interference may be especially valuable.

How experience plays out in practice at the middle school level is an empirical question. On the one hand, experience may be particularly useful at this level because of the trials of teaching adolescent students. On the other hand, the challenges of teaching at the middle school level may lead to teacher burnout that could reduce the effectiveness of teachers who remain for extended periods of time. Consistent with these or other considerations, teacher turnover rates are higher in middle schools than in other levels of schooling (Clotfelter et al. 2007; Boyd et al. 2008; Marinell and Coca 2013).

3. DATA AND MODELS

We rely on administrative data on teachers and students in North Carolina from the North Carolina Education Research Data Center. As we noted earlier, we define students in middle schools as those who are in grades 6, 7, or 8 regardless of the grade configuration of their schools. About 86 percent of such students in North Carolina are in schools serving grades 6–8 and the other 14 percent are in schools with other grade configurations, including but not limited to grades K–6 or 7–12. We work with around 250,000 students each year, for a total of about 1.2 million student-year observations. These data are available for the period 2006–2011, which (given the models include lagged variables) permits us to work with the academic years 2006–07 to 2010–11. Prior to that time, it was not possible to link students to their teachers except through information on the proctors of the end-of-course exams. For elementary students the proctor was typically the teacher in a self-contained classroom and was hence the relevant teacher for the analysis. For middle school students, that was often not the case.[1] Starting in the 2005–06 school year, the state has provided information on the teachers of each student by subject. In addition to permitting new analysis of teacher effectiveness at the middle school level, these data provide stronger teacher–subject–student links than at the elementary school level, where teachers are more frequently linked to the wrong subject (Harris and Sass 2011).

Nonetheless, we must still determine the teachers who are likely to be most responsible for a student's performance on the end-of-grade tests in reading and math and a student's non-test behaviors. At the middle school level, students typically have some choice of classes and do not all enroll in the same set of courses. In addition, a few students may enroll in more than one math or language arts course. As a result, the task of determining the main teacher is more complicated than at the elementary school level, where often a single teacher in a self-contained classroom teaches both math and reading. It is also more complicated than at the high school level in North Carolina where end-of-course tests are more directly linked to courses and the teachers who teach them. At the middle school level, our strategy for student–teacher matching first identifies each student's math and ELA classroom with the highest course membership that year, and then matches the student to the primary teacher in each of those classrooms. For example, if an eighth grader takes the standard eighth grade math course (determined as the course with the highest proportion of eighth grade students), we assign as the student's main math teacher the teacher of that student in that course. If a student does not take that course, we look for the next most enrolled math course that the student is enrolled in and use the teacher of that course as the main teacher. We follow a similar procedure for ELA.[2] For each relevant teacher defined in this way, we can compute from the administrative records years of experience and credentials, such as licensure test scores, National Board Certification, and competitiveness of the teacher's undergraduate institution.

At the student level, we have student test scores in math and ELA that can be matched over time. In addition, we make use of student-level information on absenteeism, disciplinary offenses, time spent on homework, and free time spent reading. Details about these student behavior measures appear in section 5.

Our goal is to determine how one characteristic of teachers, namely, their years of experience, affects their ability to improve their students' test scores or to change their behavior in desirable ways. To do so, we estimate a variety of models at the level of the individual student. To simplify the presentation at this point, we focus here on test scores alone, leaving for later the modeling of student behaviors. Although we estimate separate models for student performance in math and ELA, we simplify here

1. In a previous Clotfelter, Ladd, and Vigdor effort to match students to teachers in middle school, only 34 percent of sixth grade students, 29 percent of seventh grade students, and 26 percent of eighth grade students could be matched to their math teachers (Clotfelter, Ladd, and Vigdor 2007b). Although the match rates were far higher for elementary school teachers, recent research has documented the potential for mismatch of students to teachers at that level as well (Isenberg, Teh, and Walsh 2015).

2. For math, 81.3 percent of the students are in math (code 2001), 8.5 percent are in algebra 1 (code 2023), 4.7 percent are in introductory mathematics (code 2020), 3.3 percent are in accelerated middle school math (code 2003), and smaller percentages are in a variety of other courses. For language arts, 90.4 percent of the students are in language arts (code 1010), 6.1 percent are in reading (code 1001), 1.0 percent are in English as Second Language (code 1038), and smaller percentages are in a variety of other courses.

Table 1. Proportions of Groups of Students Who Have Teachers with Particular Credentials: Seventh Grade, 2010

	Experience 0–1 Years	Experience 2–5 Years	Uncompetitive College	Test Score < -1 SD	Without Advanced Degree	Without National Board Certification
Math Teachers						
By race of student						
Black	0.13	0.25	0.29	0.16	0.70	0.94
Hispanic	0.13	0.24	0.23	0.10	0.70	0.91
White	0.09	0.20	0.22	0.07	0.65	0.86
By eligibility for free lunch						
Free lunch eligible	0.12	0.23	0.27	0.12	0.69	0.91
Not free lunch eligible	0.09	0.21	0.21	0.08	0.65	0.86
English/Language Arts Teachers						
By race of student						
Black	0.10	0.28	0.23	0.14	0.64	0.90
Hispanic	0.06	0.23	0.18	0.06	0.62	0.84
White	0.08	0.27	0.19	0.09	0.63	0.86
By eligibility for free lunch						
Free lunch eligible	0.08	0.27	0.22	0.10	0.64	0.88
Not free lunch eligible	0.07	0.23	0.17	0.06	0.60	0.84

by referring to student achievement (A) as the outcome variable, without attention to subject. Two primary modeling challenges arise. One relates to the nonrandom distribution of students to teachers, and the other relates to the specification of the experience variables.

Nonrandom Distribution of Students to Teachers

As has been well documented in the literature, students are not randomly distributed among teachers, with more advantaged students typically found in classrooms with more qualified teachers (Betts, Rueben, and Danenberg 2000; Lankford, Loeb, and Wyckoff 2002; Clotfelter et al. 2007). Table 1 illustrates this pattern for our North Carolina data. Although the same pattern holds for each middle school grade and for each year of our data, we illustrate it for teachers of seventh graders in 2010, with math teachers in the top panel and ELA teachers in the bottom panel. The students are grouped by race and family income, as measured by their eligibility for free or reduced-price lunch (FRPL). The columns refer to characteristics of the students' teachers, with the entries representing the average proportions of teachers served by each specified group of students. In all cases, the teacher qualifications are defined so that higher proportions represent teachers with weaker qualifications. For example, moving down the first two columns we see that, compared with white students (as well as in most cases Hispanic students), black students have higher proportions of math teachers with limited experience (defined either as 0–1 or 2–5 years). In particular, the probability that a black seventh grader has a teacher with five or fewer years of experience is 38 percent, in contrast with 29 percent for a white student. Similarly, students who are eligible for FRPL also have higher proportions of inexperienced teachers than their counterparts from more advantaged families. Moving across the columns shows that similar patterns emerge for other teacher characteristics, such as whether the teacher graduated from an uncompetitive college, has a licensure test score more than one standard deviation below the mean, does not have an advanced degree, or is not National Board–certified.

These distributional patterns are of policy interest in their own right. To the extent that teacher qualifications are causally linked to student achievement or other outcomes (as studies have shown a number of them to be), an uneven distribution of students of the type shown in the table translates into uneven outcomes that work to the detriment of disadvantaged students. In addition, this nonrandom sorting is directly relevant for efforts to model how teachers affect student achievement and other outcomes. The concern is that any estimates of teacher effectiveness may be biased upward by the fact that the teachers with stronger qualifications are more likely to be teaching the more advantaged students.

The standard approach for addressing this challenge is to control statistically as fully as possible for the characteristics of the students. Following the literature (see, e.g., Clotfelter, Ladd, and Vigdor 2007a, b and Wiswall 2013), we use two approaches for doing so. In model 1 we control for the student's prior year achievement and a number of time-invariant characteristics of students, such as race or eligibility for FRPL,[3] and time-varying characteristics, such as whether they are new to the school. In model 2, we replace the lagged achievement variable and the time-invariant student characteristics with student fixed effects. These fixed effects capture both the observable and unobservable characteristics of the students that affect achievement, and might be viewed as a measure of the student's ability and motivation. In both models we also include a number of classroom level variables to separate any teacher effects from classroom effects, and in model 1 we also include school fixed effects, which means we are identifying the teacher effects based on the within-school variation in teacher experience.

Model 1 takes the form

$$A_{ijsgt} = \beta_0 + \beta_1 A_{i,t-1} + \beta_2 TQ_{jt} + \beta_3 X_i + \beta_4 X_{it} + \beta_5 C_{it} + \theta_j + \delta_s + \pi_{gt} + \mu_{ijsgt}, \tag{1}$$

and model 2 takes the form

$$A_{ijsgt} = \gamma_0 + \gamma_1 TQ_{jt} + \gamma_2 X_{it} + \gamma_3 C_{it} + \theta_j + \alpha_i + \pi_{gt} + \mu_{ijsgt}, \tag{2}$$

where A_{ijsgt} is student achievement, for student i matched to teacher j, in school s, grade g, and year t, and $A_{i,t-1}$ is lagged achievement. All the other variables are vectors. The vector $\mathbf{TQ}_{jt}$ represents time-varying characteristics of teachers, such as their years of experience. As we explain subsequently, the time-invariant characteristics are captured by teacher fixed effects (θ_j).[4] The vector $\mathbf{X}_i$ in model 1 refers to time-invariant student characteristics and $\mathbf{X}_{it}$ in both models refers to student characteristics that are time-varying, such as whether the student has moved to a new school. The vector $\mathbf{C}_{it}$ refers to

3. In practice, eligibility for FRPL may not be time-invariant because of possible changes in the income level of the student's parents over time or because the student may choose not to participate in the program perhaps because of concerns about being stigmatized as poor. The fact that we are missing some FRPL data for some students for some years has forced us to define a time-invariant variable based on whether we ever observe the child on FRPL in our sample.
4. If we were not including teacher fixed effects, this vector would include variables such as the teacher's licensure test score, National Board certification, advanced degree, and competitiveness of her undergraduate institution.

classroom characteristics such as class size and composition of peers, including their average prior year achievement.[5] In addition, both models include grade-by-year fixed effects (π_{gt}), model 1 includes school fixed effects (δ_s), and model 2 includes student fixed effects (α_i). The grade-by-year effects serve to control for differences across grades that may also differ by year, the school fixed effects for differences across schools, and the student fixed effects for observable and unobservable time-invariant aspects of student ability. The final term, u_{ijsgt}, is an error term for the *i*th student.[6]

Neither approach is perfect. In the model 1 specification, the lagged achievement variable serves as a proxy for all the prior inputs, including those at the family, school, and community levels, that influenced student achievement prior to year *t*. Although the inclusion of lagged achievement makes intuitive sense (in that one would not want to attribute to a student's seventh grade teacher, for example, all the student's achievement up to that time), this specification requires some significant assumptions about linearity and constancy of decay that may not all be met in practice.[7] Within such a model, the coefficient β_1 on the lagged term is typically interpreted as the share of knowledge that persists from one year to the next, or alternatively $1 - \beta_1$ is the rate at which knowledge decays. If the error terms are serially correlated over time as they are likely to be, however, A_{t-1} is not fully exogenous.[8] Moreover, measurement error poses another statistical problem, albeit one we are able to address in our modeling effort. Nonetheless, equation 1 has become the standard starting point for estimating the effects of teacher credentials on student achievement.[9]

The model 2 specification with student fixed effects excludes the lagged achievement term. Although, technically, lagged achievement could be included, it is not desirable in our sample not only because it would be endogenous but also because of the limited number of observations for each student. As shown in a technical note by Rivkin

5. Jackson (2012a,b) argues that at the high school level it is also important to control for a student's track because more experienced teachers may sort into tracks with unobservably higher achieving students or because tracks with more experienced teachers may provide other forms of supports. Those arguments are far less compelling for this study of middle school teachers, because while tracking may still occur in middle schools it is likely to be specific to a subject such as math rather than applying to a large set of courses as in, for example, the college track at the high school level. We follow the recommendation of Protik et al. (2013) and include as a control variable the mean lagged average achievement in the teacher's class as well as other class characteristics. We note as well that the inclusion of student fixed effects in model 2 obviates Jackson's first concern.
6. In an earlier version of this paper, we also included results for models in which the dependent variable was expressed as the gain in test scores. That approach solves the endogeneity problem associated with having the lagged achievement variable on the right hand side of the equation, but at a big cost. In effect it constrains the coefficient on the lagged term to be 1, which would imply that there is no decay in learning from one year to the next.
7. One assumption is that the effect of teacher credentials in the contemporaneous year is constant across years and the relationship is linear. The second is that any decay in student achievement, or knowledge, from one year to the next occurs at a constant rate. As a result, the rate at which a student's knowledge persists from one year to the next occurs at a constant rate, designated by the coefficient β_1 in the model (Clotfelter, Ladd, and Vigdor 2007b; Ladd 2008).
8. Todd and Wolpin (2003) emphasize that unless the student ability endowment decays at the same rate as the input effects, the lagged term representing baseline achievement will be correlated with the error term and the coefficients cannot be consistently estimated with ordinary least squares. Moreover, the endogeneity of the baseline achievement will affect not only the coefficient of the lagged term but also estimates of all the input effects.
9. We also estimated alternative model 1 regressions that included cross-subject lagged test scores to more fully account for prior student achievement. The findings related to the teacher experience variables were not affected.

(2006), the coefficients of interest that emerge from this model 2 specification are likely to be slightly downward biased.

Specification of the Teacher Experience Variables

The starting point for specifying the teacher experience variables is the inclusion of teacher fixed effects in both models. They are needed so that we can separate the contribution of additional years of experience to teachers' effectiveness from time-invariant measures of teacher quality. As noted by Wiswall (2013), the problem is in part one of selection: The teachers who choose to remain in teaching may differ in terms of quality from those who choose to leave. In addition, teachers of different cohorts may be of different quality. By including teacher fixed effects, we are in effect holding teacher quality constant as we focus on the returns to additional years of teaching.

Once we control for teacher fixed effects, however, the year variable (or in our case the grade-by-year variable) which is intended to control for state-wide changes in policy or tests that affect student achievement is perfectly correlated with the experience variable for the typical teacher who remains in the same grade. Researchers have approached this problem in various ways. One is to identify the experience effects from the performances of the teachers with discontinuous career paths for whom the departures and returns break the correlation. Papay and Kraft (2015) correctly criticize such an approach on the grounds that such teachers are atypical and clearly not representative of all teachers. Another far more common approach is to group the teacher experience variables into bins (e.g., 1–2, 3–5, 6–10, >10 years) and to estimate an average experience effect for each of the bins. With this approach, the year effects are identified by variation within each of the bins, and then those year effects are, in effect, used in the estimation of the returns to experience. Underlying this approach is the assumption that the returns to experience are similar within each of the bins. If the returns are rising, as they undoubtedly would be during the early years of a teacher's career (which are included in the initial bins), the year effects will be overestimated, which leaves too little of the remaining variation to show up as experience effects. Hence, that approach is likely to lead to severe underestimates of the returns to experience. A more sophisticated version of this model uses individual indicator variables for the early years of experience, say through year 10, and then simply constrains the experience effect to be zero for later years (see Rockoff 2004). That approach is undoubtedly superior to the standard bins approach but imposes an assumption of no returns after year 10 that might well not be correct.

Papay and Kraft (2015) have proposed an alternative two-stage approach. They recommend first estimating year effects from a model that does not include teacher fixed effects. Then, in the second stage, they estimate the experience effects within a model in which the year effects are constrained to those estimated in the first stage. The identifying assumption of this approach is that the returns to teacher experience do not differ across cohorts of teachers within the years of the sample. That is, a teacher with ten years of experience in the first year of the sample (in our case 2007) and a teacher with ten years of experience in the last year of our sample (here, 2011) would be equally effective on average. If, for a given amount of experience, the more recent cohorts of teachers are more effective than the earlier cohorts, on average (perhaps because of rising state standards for teachers, a rising supply of qualified teachers, or differentially high attrition of higher quality teachers), the estimated experience effects will be downward biased. As we document later, we find that at specified experience levels, the recent cohorts in our data are indeed more effective than earlier cohorts and this is true despite differentially low attrition rates of the higher quality teachers within a single cohort. Hence the identifying assumption of the two-stage approach is not met in our data. Nonetheless, we present one set of results using Papay and Kraft's two-stage approach for purposes of comparison.

Our preferred specifications are those with indicator variables for experience for each of the years 1–12, with the higher experience levels grouped into bins. Given prior evidence and the findings we present herein that teachers improve at a more rapid rate during their early years, it is far preferable to use bins for the latter years than for the early years. This procedure minimizes the bias in the estimated year effects and permits us to estimate a less biased set of experience effects. As we note later, the main patterns are not very sensitive to the number of indicator variables we include.

4. ACHIEVEMENT RESULTS

The main achievement results for the experience variables for both math and ELA teachers are reported in table 2. For each subject the table includes results from the lagged achievement specification (model 1), the student fixed effect specification (model 2), and the two-stage approach (based on the model 2 specification). All models include teacher fixed effects. Figure 1 shows these results graphically for our preferred model 2 specifications.

Model 1 includes controls for a number of student characteristics, including race, parental education, FRPL eligibility, special status (such as limited English, gifted, and special needs), and grade retention. In addition, it includes the student's lagged achievement. We report model 1 estimates derived from an errors-in-variables regression procedure to correct for measurement error in the lagged achievement variable.[10] All the models include two other time-varying student variables: (1) an indicator that the student changed schools and (2) an indicator that the student made a structural school change, such as a move from an elementary school to a middle school. In addition, they all include indicators for whether the teacher is the same race or gender as the student. Finally all the models include a large set of classroom characteristics, including a series of class size indicators and measures reflecting the composition of the class. These variables help us distinguish the effects of teachers from the effects of the classrooms in which they teach. The full models are included in Appendix table A.1.

The entries in the table are the contributions of experience to student test scores. Because we have normalized the test scores to have a mean of 0 and standard deviation of 1, each of the entries refers to fractions of a standard deviation. Although a coefficient of 0.06, for example, which represents six-hundredths of a standard deviation, may appear small, we argue in the following that it is large enough to be meaningful

10. From the North Carolina Department of Public Instruction (NCDPI 2014), we know typical reliability scores for reading and math End of Grade assessments (0.89 for reading and 0.92 for math). With this information, we implement the *eivreg* command in Stata to correct for measurement error. This procedure requires first demeaning variables by teacher and school, then estimating the errors-in-variables regression, and finally correcting the standard errors with a degrees of freedom adjustment.

Table 2. Returns to Experience for Math and ELA Teachers (Models 1, 2, and Two-Stage)

	Math			ELA		
Teacher experience	**Model 1[a]**	**Model 2**	**2-Stage**	**Model 1[a]**	**Model 2**	**2-Stage**
No experience (base)						
Experience 1 year	0.0662** (0.004)	0.0708** (0.004)	0.0695** (0.004)	0.0246** (0.004)	0.0206** (0.005)	0.0181** (0.004)
Experience 2 years	0.0917** (0.004)	0.0909** (0.005)	0.0891** (0.005)	0.0312** (0.004)	0.0277** (0.005)	0.0272** (0.005)
Experience 3 years	0.1013** (0.005)	0.0973** (0.006)	0.0956** (0.005)	0.0373** (0.005)	0.0361** (0.006)	0.0380** (0.006)
Experience 4 years	0.1151** (0.005)	0.1104** (0.007)	0.1070** (0.006)	0.0389** (0.005)	0.0341** (0.007)	0.0374** (0.007)
Experience 5 years	0.1372** (0.006)	0.1290** (0.007)	0.1280** (0.007)	0.0402** (0.006)	0.0395** (0.008)	0.0454** (0.008)
Experience 6 years	0.1406** (0.006)	0.1405** (0.008)	0.1374** (0.008)	0.0422** (0.007)	0.0426** (0.009)	0.0499** (0.008)
Experience 7 years	0.1530** (0.007)	0.1513** (0.009)	0.1492** (0.008)	0.0494** (0.007)	0.0494** (0.010)	0.0567** (0.009)
Experience 8 years	0.1610** (0.007)	0.1502** (0.010)	0.1466** (0.009)	0.0512** (0.008)	0.0528** (0.011)	0.0623** (0.010)
Experience 9 years	0.1567** (0.008)	0.1471** (0.011)	0.1431** (0.010)	0.0520** (0.009)	0.0596** (0.012)	0.0686** (0.011)
Experience 10 years	0.1621** (0.009)	0.1570** (0.011)	0.1516** (0.010)	0.0665** (0.009)	0.0671** (0.013)	0.0790** (0.011)
Experience 11 years	0.1688** (0.009)	0.1694** (0.012)	0.1637** (0.011)	0.0698** (0.010)	0.0722** (0.014)	0.0856** (0.012)
Experience 12 years	0.1766** (0.010)	0.1770** (0.013)	0.1695** (0.012)	0.0726** (0.011)	0.0779** (0.015)	0.0915** (0.013)
Experience 13–20 years	0.1716** (0.011)	0.1752** (0.014)	0.1679** (0.013)	0.0801** (0.012)	0.0765** (0.016)	0.0920** (0.014)
Experience 21–27 years	0.1848** (0.012)	0.1701** (0.016)	0.1621** (0.014)	0.0825** (0.014)	0.0789** (0.018)	0.1009** (0.016)
Experience 28+ years	0.1711** (0.014)	0.1545** (0.018)	0.1468** (0.016)	0.0792** (0.016)	0.0790** (0.021)	0.1042** (0.019)
Year × Grade FE	YES	YES	YES	YES	YES	YES
School FE	YES	NO	NO	YES	NO	NO
Student FE	NO	YES	YES	NO	YES	YES
Teacher FE	YES	YES	YES	YES	YES	YES
Observations	1,237,088	1,237,088	1,322,296	1,241,452	1,241,452	1,316,972
R^2	0.712	0.936	0.947	0.689	0.925	0.951

Notes: Robust standard errors in parentheses; refer to table A.1.

[a]Model 1 estimates use errors-in-variables regression to correct for measurement error of the lagged test score.

$^{**}p < 0.01$.

and policy-relevant. Given that the coefficients increase with years of experience, we conclude that, on average, teachers become more productive as they gain experience.

Two points are worth noting. First, the subject-specific returns to experience are quite similar across the three specifications. In most cases, however, and especially for math, the two-stage results (in columns 3 and 6) are somewhat smaller than those from the comparable single-stage model (in columns 2 and 5). That pattern is consistent with our finding that the key assumption underlying the two-stage approach is not met, especially for math teachers. In particular, we find that, when controlling for years of experience, teachers in the older cohorts in our sample (e.g., those with ten years of

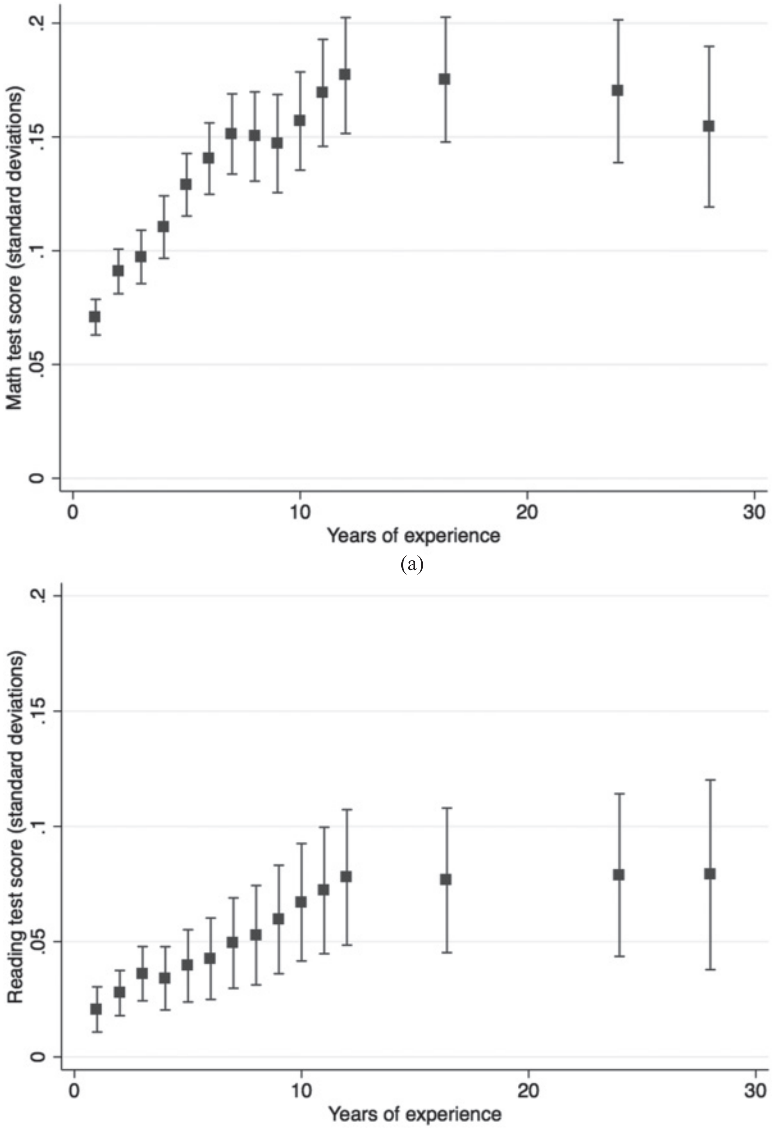

Figure 1. Test Score Results. A: Math. B: Reading.

experience in 2007) are somewhat less effective than those in the younger cohorts (e.g., those who do not have ten years of experience until 2011), where effectiveness is measured by coefficients on the teacher fixed effects.[11] As a consequence, the first stage

11. For the majority of teachers of a given experience level, fixed effect estimates were larger on average in 2011 than 2007 ($p < 0.01$). This group includes: math teachers with five years of experience (0.004 in 2011 to −0.001 in 2007), ELA teachers with ten years of experience (0.006 to −0.006), math teachers with ten years of experience (0.005 to −0.011), and math teachers with twenty years of experience (0.026 to −0.002). For one group—ELA teachers with five years of experience—fixed effect estimates were instead larger on average in 2007 than in 2011 (–0.005 to 0.010, $p < 0.01$). For ELA teachers with twenty years of experience, we cannot reject the null hypothesis that fixed effect estimates are equal between 2007 and 2011, an outcome that may

of the two-stage model attributes too much of the achievement to the year variables, which in turn leads to a small downward bias in the experience coefficients. We have no strong reason to prefer either of the two basic models over the other, but have a slight preference for model 2, both because it avoids concerns about the endogeneity of the lagged achievement term in model 1 and because that specification makes most sense for the non-test score models we present subsequently. The reader should interpret the model 2 results as conservative, however, given that those models exclude the lagged achievement term.

Second, the returns to experience differ in magnitude across the two subjects. Consider the model 2 estimates for math teachers. Consistent with prior research, the largest returns to experience appear in the early years. Specifically, one year of experience enables a teacher to raise student achievement by about 0.07 standard deviations more than having no experience. Importantly, however, that typical teacher continues to become more productive as they gain more experience, although at a lower rate. By their fifth year, their productivity is 0.13 standard deviations higher than with no experience, and continues to rise until a peak of close to 0.18 standard deviations after twelve years. At that point, their productivity levels off, but does not decline until after twenty-eight years of experience. Figure 1a portrays this rising pattern graphically, with confidence intervals around each estimate.

From the model 2 estimates for ELA teachers, we see the initial jump in productivity for teachers with one year of experience is far smaller than for math teachers at about 0.02 standard deviations. One potential explanation for this smaller coefficient is that the skill needed to teach middle school English at a minimal level may be less than the skill needed to teach middle school math. Therefore, ELA teachers with no experience are, in effect, more effective than math teachers with no experience. A related possibility is that learning how to teach English well at the middle school level may be easier than learning how to teach math, given the less technical nature of the ELA material. A third is that teachers may have less effect on verbal test scores than on math test scores because of the larger contribution of nonschool factors, such as family. In any case, the pattern suggests that the typical ELA teacher, like the math teacher, becomes more productive over time. In this case, their productivity continues to increase monotonically at least through twelve years of experience, after which time it levels off, at least according to model 2. After about twenty-five years of experience, a teacher of given quality is about 0.04 standard deviations more effective in raising test scores than they were when they had five years of experience. The pattern is portrayed in figure 1b.[12]

Formal Wald tests confirm our graphical evidence that the effect of twelve years of experience is significantly greater than the effect of four years of experience for both math and ELA teachers. That is, teachers continue to improve well after their first few years of teaching.

be due to small sample size. These test results suggest that the experience results for math teachers in the two-stage model are likely to be downward biased but the direction of bias for ELA teachers is somewhat less clear.

12. Results from alternative experience variable specifications were consistent with our main results. Model 1 and model 2 regressions with single indicators for the first twenty years of teaching experience, instead of only the first twelve, and then bins after twenty-one years, generated nearly identical results to those presented in table 2.

One might be concerned about potential bias in our estimated returns to experience given that different cohorts of teachers may have different average abilities and our estimates are based on only five years of data. We note, however, that differences in average cohort quality will not bias our estimates as long as the returns to experience do not differ across cohorts. This is true because the teacher fixed effects adjust for the quality differences. Thus, the observation (documented in the following) that the younger teachers in our sample are of higher quality than the older teachers is not a problem, as long as the average returns to experience do not differ across cohorts.[13] Similarly, the use of teacher fixed effects obviates the need to adjust for attrition rates that differ by teacher quality under the same condition, namely, that average returns to experience are similar among the stayers and the leavers.[14]

Potentially more problematic would be differences by cohort (or between stayers and leavers) in the returns to experience, although it is difficult to predict the direction of the bias. If the returns to experience are higher for the younger cohorts, our estimates will overestimate the returns in the early years because the younger cohorts are overrepresented in the set of teachers used to estimate the returns in the early years, and will underestimate returns in the later years. Correspondingly, if the returns to experience are higher for the older cohorts, our estimates will underestimate the returns to experience in the early years because the older cohorts are underrepresented and will overestimate the returns in the later years. Although such biases are possible, statistical tests (described in Appendix table A.2) based on our short five-year sample provide no indication that they are likely to be large.[15] In principle, one could avoid biases of this type by estimating the models for a single cohort of teachers. Given our five-year sample window, however, restricting the sample to a single cohort reduces the typical sample size from over 1.2 million observations to about 25,000 observations, and does not provide the variability we need to estimate reasonable models.

Comparisons to (Mis)specifications of the Teacher Variables

Importantly, our estimates need not mean that a typical North Carolina middle school teacher with many years of experience is far more effective than a typical teacher with less experience. The reason is that by including teacher fixed effects in our models, we are in effect inferring returns to experience for individual teachers, and are thereby, in effect, holding constant the teacher's time invariant characteristics such as their

13. We have confirmed this assertion by running simple simulations. These simulations are based on 600 teachers for whom we specify the same true experience profile but different average quality levels by cohort. In contrast to regressions based on models without teacher fixed effects, which generate biased estimates of the true experience profile, models with teacher fixed effects generate the correct estimates.
14. Once again, simple simulations confirm that even extreme levels of differential attrition do not bias the estimates of returns to experience. Consistent with the patterns in our actual data, we set higher departure rates for teachers of lower quality than those with higher quality. Specifically, the chances of departing after each year teaching were 0.05 percent in the top 10th percentile of ability, 0.5 percent in the top 25th to 10th percentile, 2.5 percent in the 75th to 25th percentile, 12.5 percent in the bottom 25th to 10th percentile, and 25 percent in the bottom 10th percentile. Estimates with teacher fixed effects were statistically indistinguishable from true returns to experience.
15. These tests allow for the selected experience effects (those for years 5, 7, 9, and 11) to differ between the early and late cohorts. In table A.2 the coefficients of the interaction terms are very small and only one is even marginally significant. These coefficients provide no support for the conclusion that returns to experience differ by cohort across a range of experience levels, although we fully acknowledge that the test applies only to comparisons of returns to experience among adjacent cohorts.

Table 3. Returns to Experience for Math and ELA Teachers: Alternative Specifications (Model 2)

	No Teacher Fixed Effects			Bins of Experience Level	
Teacher Experience	Math (1)	ELA (2)	Teacher Experience	Math (3)	ELA (4)
No experience (base)					
1 year	0.0608** (0.003)	0.0199** (0.003)	1–2 years	0.0667** (0.004)	0.0166** (0.004)
2 years	0.0630** (0.004)	0.0284** (0.003)			
3 years	0.0648** (0.003)	0.0320** (0.004)	3–5 years	0.0748** (0.005)	0.0213** (0.005)
4 years	0.0613** (0.004)	0.0251** (0.004)			
5 years	0.0670** (0.004)	0.0296** (0.004)			
6 years	0.0742** (0.004)	0.0291** (0.004)	6–12 years	0.0910** (0.006)	0.0234** (0.006)
7 years	0.0776** (0.004)	0.0326** (0.004)			
8 years	0.0631** (0.004)	0.0325** (0.004)			
9 years	0.0577** (0.004)	0.0306** (0.004)			
10 years	0.0593** (0.004)	0.0334** (0.004)			
11 years	0.0678** (0.004)	0.0290** (0.004)			
12 years	0.0686** (0.004)	0.0314** (0.004)			
13–20 years	0.0710** (0.003)	0.0343** (0.003)	13–20 years	0.0851** (0.006)	0.0170* (0.007)
21–27 years	0.0750** (0.003)	0.0361** (0.003)	21–27 years	0.0730** (0.007)	0.0125*** (0.008)
28-plus years	0.0645** (0.003)	0.0378** (0.003)	28-plus years	0.0576** (0.005)	0.0086 (0.006)
Year × Grade FE	YES	YES		YES	YES
School FE	NO	NO		NO	NO
Student FE	YES	YES		YES	YES
Teacher FE	NO	NO		YES	YES
Observations	1,206,749	1,215,900		1,281,770	1,281,412
R^2	0.019	0.015		0.923	0.935

Notes: Robust standard errors in parentheses.

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.1$.

basic ability and motivation. If it were the case that the average basic ability (and/or motivation) of teachers differed across cohorts of teachers grouped by their years of experience, the patterns generated from models without teacher fixed effects could potentially differ quite substantially from those generated by our preferred models. The results reported in columns 1 and 2 of table 3, which come from models with no teacher fixed effects, illustrate this is indeed the case. Although these alternative models include a number of teacher-specific characteristics and credentials, they do not fully control for teacher ability and motivation. The differences between these reported patterns and those from our preferred models are quite large. Without the teacher fixed effects, the coefficients for math teachers increase from 0.061 to a peak of 0.077, and for ELA teachers they increase from 0.020 to a peak of 0.033. These increases are far smaller than the comparable increases from 0.071 to 0.170 for math and from 0.021 to 0.079 for ELA that emerge from our preferred models with teacher fixed effects.

The differences in the estimated experience trajectories are attributable to the fact that low-ability teachers are overrepresented among the more experienced teachers within our sample. At first, this pattern might seem surprising in light of various studies showing that less-effective teachers (as measured by the gains in their students' test scores) are more likely to leave the profession than more-effective teachers during the early years of their careers (Hanushek et al. 2005; Krieg 2006; Boyd et al. 2008; Goldhaber, Gross, and Player 2011). Indeed, that departure pattern is evident in our data as well. Of the group of new ELA teachers entering in 2007, those who remain in our data in 2011 have fixed effect estimates 0.036 standard deviations higher than those who leave during this time period. For math teachers, the gap between those who stay and those who leave is 0.073 standard deviations.

But something far more important is driving the experience patterns—namely, despite these differential departure rates, the more experienced cohorts of teachers have less basic ability than the more-recent cohorts. A snapshot of ELA teachers teaching during the 2008–09 school year shows that average fixed effect coefficients are 0.019 for teachers with fewer than five years of experience, −0.005 for teachers with five to fifteen years of experience, and −0.020 for teachers with over fifteen years of experience. This pattern appears for math teachers as well, with average fixed effect coefficients of 0.045 for teachers with fewer than five years of experience, −0.025 for teachers with five to fifteen years of experience, and −0.023 for teachers with over fifteen years of experience. Had the departure rates not differed by teacher quality, that pattern would have been even stronger. Possible explanations for this pattern could include general labor market trends as well as the fact that over the relevant period, North Carolina raised the standards for a growing number of teachers.

Finally, we confirm that the specification of the experience variables themselves matters and, in particular, the common practice of putting the experience variables into categories or bins rather than including indicators for each level of experience leads to seriously downward biased estimates of the returns to experience. The results in columns 3 and 4 of table 3 show returns to experience when experience is specified by a series of bins: 1–2 years, 3–5 years, 6–12 years, 13–20 years, 21–27 years, and 28-plus years. As we noted earlier, any significant increases in teacher productivity *within* these bins would be picked up by the grade-by-year fixed effects, rather than by the experience variables, thereby leading to downward biased estimates of the returns to experience. As seen in table 3, coefficients for math teachers rise from 0.0667 for 1–2 years of experience to a peak of 0.0910 for 6–12 years of experience. For ELA teachers, the coefficients rise from only 0.0166 for 1–2 years of experience to 0.0234 for 6–12 years of experience. As predicted, these estimated returns to experience are much smaller than in our preferred model specification with more detailed experience variables.

The bottom line is that many prior studies may bias the test-based returns to experience either because they do not control for teacher fixed effects or they incorrectly specify the experience variables, or both.

Table 4. Coefficients of Student-Level Variables from Model 1

Student Variables	Math	ELA
Male	−0.0102**	−0.0020
	(0.001)	(0.001)
Black	−0.0320**	−0.0681**
	(0.002)	(0.002)
Hispanic	0.0301**	0.0075**
	(0.002)	(0.002)
Other race	0.0393**	0.0029
	(0.002)	(0.002)
Age in grade 3	−0.0499**	−0.0437**
	(0.001)	(0.001)
Parent college graduate (base)		
Parent comm. college graduate	−0.0252**	−0.0201**
	(0.001)	(0.002)
Parent high school graduate	−0.0411**	−0.0436**
	(0.001)	(0.001)
Parent high school drop out	−0.0618**	−0.0665**
	(0.003)	(0.003)
Limited English	−0.0049*	−0.0613**
	(0.002)	(0.003)
Gifted in Math	0.1087**	0.0672**
	(0.002)	(0.002)
Gifted in Reading	0.0779**	0.0883**
	(0.002)	(0.002)
Special needs	−0.0486**	−0.0564**
	(0.002)	(0.002)
Subsidized lunch	−0.0437**	−0.0324**
	(0.001)	(0.001)
Repeated grade	−0.0748**	−0.0356**
	(0.003)	(0.003)
School change	−0.0319**	−0.0229**
	(0.002)	(0.002)
Structural school change	−0.0976**	−0.0858**
	(0.003)	(0.003)

Notes: These estimates use errors-in-variables regression to correct for measurement error of the lagged test score; coefficients for other variables from model 1 version of full models in table A.1.

*$p < 0.05$; **$p < 0.01$.

Interpretation of the Estimated Effects

One way to put our preferred coefficient estimates into perspective is to compare them to other coefficients. In table 4 we report results for the student-level coefficients from model 1. The signs of most of these student-level variables are consistent with our expectations based on extensive prior research on student achievement. In particular, students from less advantaged backgrounds as defined by race, education level of parents, and income all exhibit lower math and reading scores than their more advantaged counterparts. Students who are identified as gifted perform better, and those identified as having special needs, who are limited English proficient, or who are repeating the grade perform less well. In addition, students who are moving into a new school not as part of a structural change, as well as those making a structural change such as from an elementary school to a middle school, also perform less well than those not moving from one school to another. Finally, male students do about the same as female students in math but less well in ELA.

The comparison of coefficients shows that the returns to experience for math teachers are large relative to many of the student-level coefficients in the math equation. Consider, for example, the 0.129 coefficient for math teachers with five years of experience (from model 2). This estimate suggests that five years of teaching experience is more than enough to offset the −0.032 negative effect of being a black student or the −0.044 effect of being eligible for a subsidized lunch, all other factors held constant. At the same time, it is not sufficiently large to offset a combination of disadvantaging factors, such as being black, being eligible for subsidized lunch, or having a parent who is a high school dropout. In sum, the effects of teacher experience for math teachers are quite large, but experience alone may not suffice to offset all the challenges that children with multiple disadvantages bring to the classroom.

Compared with the math teachers, the returns to experience for ELA teachers are less impressive relative to the magnitudes of the student-level coefficients. The coefficient of about 0.078 for a teacher with twelve years of experience is only somewhat larger than the −0.068 coefficient for black students, and clearly not sufficient to offset the combined effects of being black and almost any other disadvantaging characteristics. In general, these patterns are consistent with earlier research that highlights the relatively stronger role that family background and home environment play on reading over math. Experience for ELA teachers is still important but other factors are even more important.

5. NON-TEST STUDENT OUTCOMES

Prior research has shown that high-quality teachers as measured by the gains in test scores of their students can generate substantial, long-term economic value for their students (Chetty, Friedman, and Rockoff 2011). Additional evidence suggests, however, that the long-term value of teachers may be significantly higher if one considers how teachers affect not just test scores, but a fuller range of non-test student behaviors. For example, Jackson (2012a) finds the causal effects of ninth grade teachers on outcomes such as absences, suspensions, grades, and on-time grade progression predict subsequent dropout and college plans above and beyond the prediction from teacher effects on test scores alone. For that reason, in this section we broaden the concept of teacher productivity beyond student test scores.

Our administrative data enable us to examine how teacher experience affects four specific measures of student behaviors: (1) number of days absent, (2) number of reported disruptive classroom offenses, (3) amount of time spent completing homework, and (4) amount of time spent reading for pleasure. Although not a comprehensive set, these behavioral measures serve as appropriate proxies for what are sometime referred to as noncognitive capabilities, known to benefit students' later education, economic, or social outcomes. For example, attendance as early as sixth grade strongly predicts the likelihood of graduating from high school (Allensworth and Easton 2007; Balfanz, Herzog, and Mac Iver 2007). Teachers who successfully encourage students to put forth effort on their homework or to manage their classroom behavior are developing self-discipline, which the literature shows does a better job of predicting adolescent academic performance than IQ (Duckworth and Seligman 2005). Academic curiosity—measured in our study by free time reading—motivates proactive and intentional learning behaviors which, in turn, result in the increased acquisition of academic and social skills (Kashdan, Rose, and Fincham 2004). And measures of student self-control, such as that required to remain respectful and focused in class, strongly predict adult health, wealth, and public safety (Moffitt et al. 2011). Moreover, the interaction of negative student behaviors such as not doing homework, skipping school, and acting out in class can worsen undesirable outcomes (Balfanz, Herzog, and Mac Iver 2007).

Measurement of Student Behaviors

Two measures appear in school administrative data: number of days absent and number of reported student offenses. We limit offenses to those recorded as disorderly conduct, inappropriate language/disrespect, insubordination, disruptive behavior, disrespect of faculty/staff, or skipping class. These offense categories represent behaviors plausibly affected by teachers and are indicative of lack of student self-control. Because the distributions of both absence and disciplinary offense counts are highly abnormal (as shown in table 6), we would prefer to use Poisson or negative binomial regression models to explain the variation in these variables. The use of high-dimensional fixed effects, however, combined with the large size of our dataset, renders these alternative regression models computationally infeasible. Therefore, we transform the absence and disciplinary offense outcomes into dichotomous variables and use linear probability models. For absences, we construct an indicator of whether or not a student has an absentee rate above the 75th percentile of the overall distribution. For offenses, we construct an indicator of whether or not the student commits any disciplinary offenses that year. As a sensitivity check, we also examine absence and offense indicators based on different thresholds: greater than 25th percentile absences, greater than median absences, greater than 90th percentile absences, greater than one offense, and greater than two offenses.

End-of-grade test forms in the North Carolina administrative dataset include questions regarding the amount of time students spend completing homework and amount of free time students devote to reading. Students are asked to choose among a number of different bins representing time spent on a weekly basis. For time spent on homework the bins are: has homework but does not do it; less than one hour; between 1 and 3 hours; more than 3 but fewer than 5 hours; between 5 and 10 hours; and more than 10 hours. For time spent reading these bins are: none; about 30 minutes; about 1 hour; between 1 and 2 hours; and more than 2 hours. We include these measures, recoded as approximate number of hours, as the final two of our non-test outcomes. Tables 5 and 6 summarize the distribution of all non-test outcome variables across the sample of students.

Each of these measures—absences, offenses, homework, and free reading—has advantages and disadvantages. The most objective and obvious measure of the four is student absences in that it tends to be reliably tracked and reported by school administrators. Moreover, it is behavior that has been most clearly linked to a student's future success, either because of its direct connection to learning (it is hard for students to learn if they do not come to class) or because of what it indicates about student motivation. One potential disadvantage of this measure is it refers to absences from

Table 5. Descriptive Summary of Categorical Student Behavior Measures

Variable	Coded No. of Hours	Frequency	Percentage
Free time spent reading			
None	0	196,814	16.6
About 30 minutes	0.5	578,513	48.9
About 1 hour	1	206,340	17.4
Between 1 and 2 hours	1.5	117,996	10.0
More than 2 hours	2	84,257	7.1
Total		*1,183,920*	*100.0*
Time spent doing homework			
No homework assigned	-	19,881	1.7
Has homework but does not do it	0	18,545	1.6
Less than one hour each week	0.5	365,368	30.8
Between 1 and 3 hours	2	523,846	44.2
More than 3 but less than 5 hours	4	157,083	13.2
Between 5 and 10 hours	7.5	86,365	7.3
More than 10 hours	10	14,533	1.2
Total		*1,185,621*	*100.0*

Table 6. Descriptive Summary of Student Behavior Measures

Variable	Statistic
Number of days absent ($N = 1{,}165{,}725$)	
25th percentile	3
Median	6
75th percentile	10
90th percentile	17
Classroom disruption offenses ($N = 1{,}475{,}402$)	
Mean	0.336
Standard deviation	1.325

school, not from a particular class. As a result, even if teachers as a group influence student absenteeism, an individual math or ELA teacher may have a more limited effect, which conceivably might not be picked up by our analysis of individual teachers.

The data on student offenses are based on voluntary teacher reports, and teachers may differ somewhat in the offenses they choose to report. We do not expect differential reporting propensities of teachers to significantly bias our estimates, however, because any teacher or administrator at the school can report offenses, and not just the math or ELA teacher matched to the student. This measure provides potentially useful information about undesirable student behaviors in school that are not captured by typical suspension records and that might be influenced by effective teachers.

The measures of homework and free reading time have the potential to provide insight into student effort and motivation outside the classroom. However, they contain two sources of measurement error. First, the variables are categorical and do not measure the time spent reading or doing homework in a continuous fashion. Second, these measures are self-reported by students. As long as the measurement error is randomly distributed across students, though, it does not pose a serious statistical problem for us because our use of the measures as outcome variables implies that any error is contained in the error terms.

Non-test Returns to Teacher Experience

Our dataset and regression models remain the same as for model 2, with a few modifications. Whereas a student's ELA and math teachers each uniquely contribute to reading and math test scores, respectively, either one of the teachers could contribute to a student's more general behaviors. However, for absences, classroom offenses, and amount of time spent completing homework, we estimate separate models for each of the ELA and math classroom samples.[16] For the amount of free time spent reading we use only the sample of students matched with ELA teachers. For absences and offenses we estimate linear probability models in which the dependent variable takes on the value 1 for absences (or offenses) that exceed a specified threshold. For time on homework and free reading, we use ordinary least squares regressions to explain variation in the dependent variables which are specified as continuous variables in fractions of hours.

Figures 2 and 3 display graphically the returns to years of teaching experience for these four outcomes based on our preferred model 2 specification, and based on the results reported in table 7 (for ELA teachers) and table 8 (for math teachers). Although the confidence intervals are very large in some cases—and much larger than those for test scores—the point estimates generally suggest positive returns from experience across a variety of student behaviors, for both ELA and math teachers.

For the absence and offense measures, both of which represent problematic student behaviors, negative coefficients that are becoming larger in magnitude would signify that teachers become more productive with experience. The coefficients on the experience variables for our absence measure represent estimated reductions in the probability that a student has a high annual absentee rate (one in the top quartile). For student offenses the coefficients measure reductions in the probability that a student has any disciplinary offense that year when offenses include disorderly conduct, inappropriate language/disrespect, insubordination, disruptive behavior, disrespect of faculty/staff, or skipping class. For homework effort and free reading, both of which are positive outcomes, if the coefficients increase with years of experience we conclude, on average, that teachers become more productive in fostering these positive student behaviors as they gain experience. For these measures, the coefficients on experience correspond to changes in the number of hours per week dedicated to homework or free reading.

16. In addition to the results reported in tables 7 and 8, we estimated an alternative set of non-test regressions using a "stacked" dataset of students matched to both their math and ELA teachers. Most students appear twice in each year of this dataset for each outcome measure. In these analyses we are in effect estimating the average effects of teacher experience across the two subjects. For absenteeism, the estimated experience coefficients from this stacked model range from −0.018 for one year of experience to −0.12 for 28 or more years of experience, which are very similar to the results reported in tables 7 and 8. The advantage of this alternative approach is that the standard errors are slightly smaller. The disadvantage is that it obscures one of the key findings from the subject-specific models of absenteeism, namely, the patterns and magnitudes are similar for math and ELA teachers.

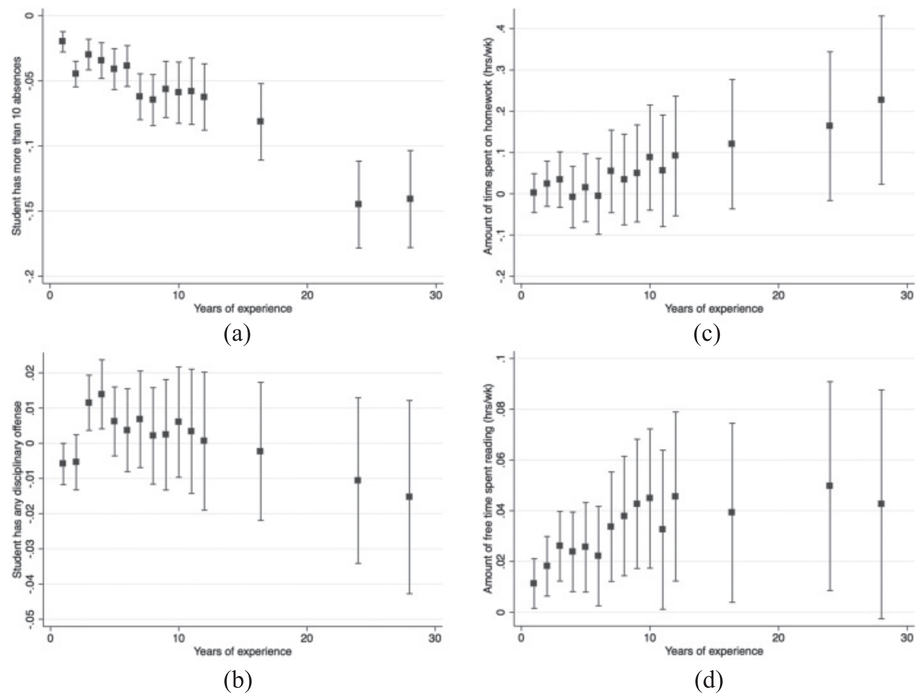

Figure 2. Non-test Student Outcome Results: ELA Teacher Sample. A: Probability of Having More Than Ten Absences. B: Probability of Having at Least One Reported Offense. C: Amount of Time Spent on Homework. D: Amount of Time Spent Reading for Enjoyment.

English Language Arts Teacher Sample

The clearest effects emerge for student absences. One year of experience enables an ELA teacher to reduce the proportion of students with high absenteeism by 2.0 percentage points and these reductions increase as they continue to gain experience. A teacher of given quality who obtains over 21 years of experience on average reduces the incidence of high student absenteeism by 14.5 percentage points (see table 7 and figure 2a). Provided what we know about the link between high absenteeism and long-term educational outcomes (Allensworth and Easton 2007; Balfanz, Herzog, and Mac Iver 2007), a reduction of 14.5 percentage points from a single school year with a more experienced teacher seems highly policy-relevant.[17] This estimate implies that, in an average classroom, replacing a new teacher with an experienced teacher could

17. Appendix tables A.5 and A.6 show results for absences and offenses with different threshold values. For absences, we replace the 75th percentile threshold of 10 absences with several other values: 25th percentile (three absences), median (six absences), and 90th percentile (seventeen absences). These results show clearly that experienced teachers are more capable of reducing absences for students with very low levels of attendance than for students with already high levels of attendance. For example, ELA and math teachers with 21–27 years of experience reduce the number of students with over three absences by 5.6 and 4.4 percent, respectively. However, ELA and math teachers with the same level of experience reduce the number of students with over seventeen absences by 18.8 and 12.2 percent, respectively. Returns to teacher experience in terms of attendance are therefore highest for higher-risk students. This pattern makes sense. Although teachers may not be able to prevent a student from getting sick, they may be able to influence attendance for students with truancy problems. For disciplinary offenses, the strongest results are for the threshold of whether students have *any* reported offenses rather than for higher number of offenses. This pattern may reflect the small number of students in any given year with multiple infractions related to disorderly conduct, inappropriate language/disrespect, insubordination, disruptive behavior, disrespect of faculty/staff, or skipping class.

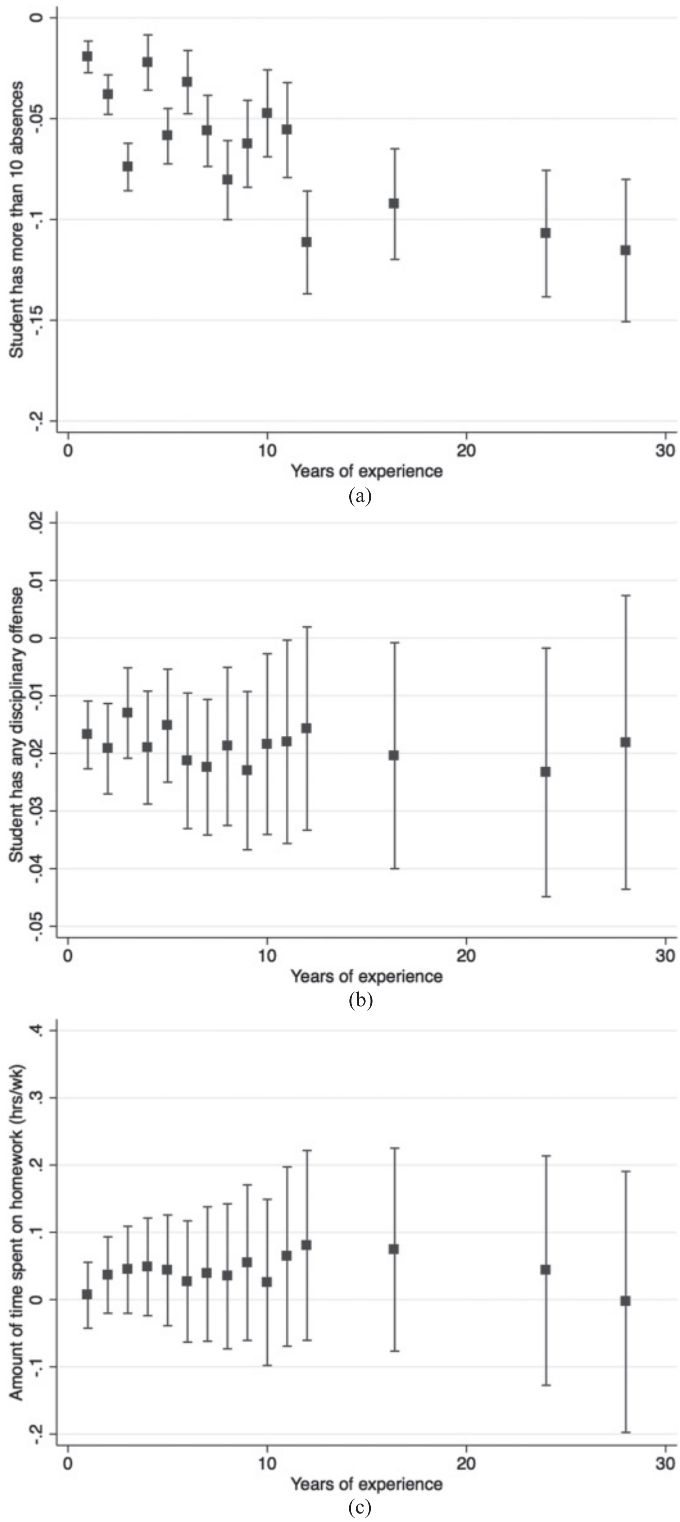

Figure 3. Non-test Outcome Results: Math Teacher Sample. A: Probability of Having More Than Ten Absences. B: Probability of Having at Least One Reported Offense. C: Amount of Time Spent on Homework.

Table 7. Returns to Teacher Experience for Non-test Student Outcomes: ELA Teacher Sample (Model 2)

	Absences > 10	Offenses > 0	Homework Effort	Free Reading
Teacher credentials				
No experience (base)				
Experience 1 year	−0.0200**	−0.0059*	0.0017	0.0113*
	(0.004)	(0.003)	(0.024)	(0.005)
Experience 2 years	−0.0448**	−0.0054	0.0243	0.0181**
	(0.005)	(0.004)	(0.028)	(0.006)
Experience 3 years	−0.0298**	0.0115**	0.0343	0.0260**
	(0.006)	(0.004)	(0.033)	(0.007)
Experience 4 years	−0.0344**	0.0139**	−0.0081	0.0238**
	(0.007)	(0.005)	(0.038)	(0.008)
Experience 5 years	−0.0410**	0.0062	0.0145	0.0256**
	(0.008)	(0.005)	(0.042)	(0.009)
Experience 6 years	−0.0386**	0.0037	−0.0063	0.0221*
	(0.008)	(0.006)	(0.047)	(0.010)
Experience 7 years	−0.0622**	0.0068	0.0543	0.0337**
	(0.009)	(0.007)	(0.051)	(0.011)
Experience 8 years	−0.0648**	0.0021	0.0345	0.0379**
	(0.010)	(0.007)	(0.056)	(0.012)
Experience 9 years	−0.0566**	0.0024	0.0494	0.0427**
	(0.011)	(0.008)	(0.060)	(0.013)
Experience 10 years	−0.0591**	0.0060	0.0876	0.0448**
	(0.012)	(0.008)	(0.065)	(0.014)
Experience 11 years	−0.0580**	0.0034	0.0557	0.0325*
	(0.013)	(0.009)	(0.069)	(0.016)
Experience 12 years	−0.0624**	0.0006	0.0916	0.0456**
	(0.013)	(0.010)	(0.074)	(0.017)
Experience 13–20 years	−0.0815**	−0.0023	0.1202	0.0392*
	(0.015)	(0.010)	(0.080)	(0.018)
Experience 21–27 years	−0.1450**	−0.0106	0.1638***	0.0497*
	(0.017)	(0.012)	(0.092)	(0.021)
Experience 28+ years	−0.1408**	−0.0153	0.2270*	0.0425***
	(0.019)	(0.014)	(0.104)	(0.023)
Grade × Year FE	**YES**	**YES**	**YES**	**YES**
School FE	**NO**	**NO**	**NO**	**NO**
Student FE	**YES**	**YES**	**YES**	**YES**
Teacher FE	**YES**	**YES**	**YES**	**YES**
Observations	1,329,206	1,332,977	976,278	980,163
R^2	0.721	0.731	0.730	0.804

Notes: Robust standard errors in parentheses; refer to table A.3.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.1$.

reduce the proportion of students with high absenteeism by more than half. It is less clear whether ELA teachers become more effective at classroom management as they gain experience—the rate of student disruptive offenses increases briefly in the third and fourth years of teaching but then decreases across higher experience levels. For the most part, these estimates for classroom offenses are not statistically significant.

Table 7 also shows that the first 20 years of experience do little to enhance ELA teacher productivity in terms of inducing students to devote more effort to homework. Although the coefficients are generally positive and increasing, they are imprecise and we cannot rule out the possibility that they reflect chance alone. The only (weakly)

Table 8. Returns to Teacher Experience for Non-test Student Outcomes: Math Teacher Sample (Model 2)

	Absences > 10	Offenses > 0	Homework Effort
Teacher credentials			
No experience (base)			
Experience 1 year	−0.0194**	−0.0168**	0.0064
	(0.004)	(0.003)	(0.025)
Experience 2 years	−0.0381**	−0.0192**	0.0364
	(0.005)	(0.004)	(0.029)
Experience 3 years	−0.0740**	−0.0130**	0.0442
	(0.006)	(0.004)	(0.033)
Experience 4 years	−0.0222**	−0.0190**	0.0486
	(0.007)	(0.005)	(0.037)
Experience 5 years	−0.0587**	−0.0152**	0.0434
	(0.007)	(0.005)	(0.042)
Experience 6 years	−0.0319**	−0.0213**	0.0269
	(0.008)	(0.006)	(0.046)
Experience 7 years	−0.0561**	−0.0224**	0.0380
	(0.009)	(0.006)	(0.051)
Experience 8 years	−0.0805**	−0.0188**	0.0345
	(0.010)	(0.007)	(0.055)
Experience 9 years	−0.0625**	−0.0230**	0.0549
	(0.011)	(0.007)	(0.059)
Experience 10 years	−0.0474**	−0.0184*	0.0255
	(0.011)	(0.008)	(0.063)
Experience 11 years	−0.0557**	−0.0180*	0.0639
	(0.012)	(0.009)	(0.068)
Experience 12 years	−0.1114**	−0.0157***	0.0805
	(0.013)	(0.009)	(0.072)
Experience 13–20 years	−0.0924**	−0.0204*	0.0742
	(0.014)	(0.010)	(0.077)
Experience 21–27 years	−0.1070**	−0.0233*	0.0431
	(0.016)	(0.011)	(0.087)
Experience 28+ years	−0.1154**	−0.0181	−0.0035
	(0.018)	(0.013)	(0.099)
Grade × Year FE	**YES**	**YES**	**YES**
School FE	**NO**	**NO**	**NO**
Student FE	**YES**	**YES**	**YES**
Teacher FE	**YES**	**YES**	**YES**
Observations	1,319,672	1,325,142	969,364
R^2	0.716	0.733	0.731

Notes: Robust standard errors in parentheses; refer to table A.4.

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.1$.

significant results emerge for teachers with more than 21 years of experience, and this suggests that very experienced teachers may promote ten- to fourteen-minute increases in the amount of time students spend on homework per week. A far clearer gradient arises with respect to free reading. As an ELA teacher's experience increases, they significantly improve the amount of free time their students spend reading, with modest but steadily increasing returns to a peak at 21–27 years of experience. The 0.050 coefficient for that experience level translates into about three extra minutes reading for pleasure weekly—about a 10 percent increase of the most frequently reported time of 30 minutes.

Math Teacher Sample

For the math teachers (table 8 and figure 3), once again substantial returns to teacher experience emerge for student absenteeism, with the absolute value of the coefficients increasing in magnitude across years of math teacher experience. Two years of math teaching experience leads to a reduction in the proportion of students with high absenteeism by 3.8 percentage points, an effect that rises to an 11.5 percentage point reduction for teachers with extensive experience. The results also suggest moderate returns for math teachers in the form of fewer students with reported disruptive offenses. By their second year of experience, a teacher reduces the number of students with disciplinary offenses by 1.9 percentage points. They continue to become more productive in classroom management until 21–27 years of experience, associated with 2.3 percent fewer students with disciplinary offense records. At best, teacher experience weakly predicts increased student effort on homework for the first four years but then the effects fade out.

In sum, the evidence shows that returns to teacher experience emerge not only in the form of higher student test scores but also in the form of some improved student behaviors. The clearest non-test score patterns emerge for student absenteeism and do so for both math and ELA teachers. For both types of teachers, Wald tests indicate the estimated coefficients for teachers with twelve years of experience exceed those for teachers with four years of experience. Similarly, more experienced ELA teachers also show some success in promoting reading for pleasure with the differences between twelve and four years again being statistically significant. The findings for student offenses are far less clear.[18] We provide coefficients from the full models for non-test outcomes in tables A.3 and A.4.

6. CONCLUSIONS AND IMPLICATIONS FOR POLICY DEBATES

The findings are quite clear. There appear to be large returns to experience for middle school teachers in the form both of higher test scores and improvements in student behavior, with the clearest behavioral effects emerging for reductions in student absenteeism. Moreover, these returns extend well beyond the first few years of teaching. The findings for student test scores challenge the conventional wisdom that teachers essentially stop improving after the first few years of teaching. Instead, teachers continue to develop long into their teaching careers. The findings related to absenteeism, and the suggestive evidence for other behaviors, show that experience can be beneficial to students in ways other than simply developing their cognitive skills. Despite our efforts to extend the analysis to student outcomes beyond test scores, we remind the reader that we are still providing at most a partial picture of the full set of potential contributions

18. One possible concern is that despite our best efforts at unbiased identification of returns to teacher experience perhaps there is some underlying relation between, for example, the experience levels of all teachers at a school with a student's behavior. As a falsification test, we regress a student's non-test outcomes on the experience level of a student's teacher in the *following* year. For six of our seven non-test outcomes, including both absences indicators, there is no significant effect of the experience of a student's teacher in time *t–1* on the student's behaviors in time *t*. Only for disciplinary offenses for ELA teachers do we find that higher teacher experience in the following year is associated with higher likelihood of a disciplinary offense in the current year. This finding may indicate some sorting of students to teachers over time, but not in the expected direction and only in the case of offenses (for which our results are not significant in any case).

teacher experience may make to the schooling of children. Other contributions would include, for example, the mentoring of new teachers and the institutional memory needed to provide a coherent school environment over time.

Importantly, our findings do not imply that a typical or average teacher with many years of experience is necessarily far more effective than a typical teacher with fewer years. The reason is that by including teacher fixed effects in our preferred models we are able to separate the returns to additional years of experience from other factors that may influence the average effectiveness of teachers at any experience level, such as differential departure rates from the sample by teacher quality or cohort effects. In our sample, the higher intrinsic average quality of the younger cohorts of teachers relative to the older cohorts means that when we remove teacher fixed effects from our models, the measured relationship between experience and gains in student test scores is sharply attenuated. This pattern arises because those models confound returns to experience with cohort effects. Even in our models with teacher fixed effects, our conclusions must be tempered by the fact that we do not observe all teachers over a full career. As we described in detail in the paper, differences in returns to experience by cohort could lead to either upward or downward biases in our estimates of the returns to experience.

Given the structure of current salary schedules, which provide higher salaries for more experienced teachers, some policy makers may be tempted to try to save money by hiring higher quality (but relatively cheap, inexperienced) teachers, and not worrying about retaining their more experienced teachers who may have been hired at a time of lower standards. Based on the estimates provided here, however, such a strategy makes sense only if two conditions hold. One is that the new cohort of inexperienced teachers has far greater intrinsic quality on average than the experienced teachers. The other is that the new high-quality teachers can be induced to remain in the profession for sufficiently long periods for the schools to benefit from the significant returns to teacher experience. If high-quality teachers do not remain in the schools, the schools face high costs both in the form of teacher turnover and in the form of the loss in productivity that comes with experience. Thus, the first imperative for schools is to recruit high-quality teachers. The challenge then is to provide working environments that support their development, and to pursue policies that explicitly recognize the value of experience in order to retain those teachers.

With its attention to some of the non-test score behaviors of students, this research also speaks to a second policy issue, the one highlighted by Jackson in his 2012a paper on the contributions of teachers to the noncognitive skills of their students. Because much of the current empirical literature on teacher effectiveness focuses on the ability of teachers to raise the test scores of their students, it is easy to lose sight of other contributions teachers make to their students. This paper provides additional support for the conclusion that teachers do more than simply raise test scores. Moreover, consistent with Jackson's findings (Jackson 2012a, p. 23) it suggests that measures of effectiveness that focus on test scores alone may understate the contribution of ELA teachers relative to math teachers. Like many other researchers, we find that experienced ELA teachers make smaller contributions to gains in test scores than experienced math teachers. At the same time, our finding that experienced ELA teachers contribute to improvements in student behaviors (such as by reducing absenteeism), at rates comparable to experienced math teachers, implies a broader measure of teacher effectiveness would better capture the relative value of ELA teachers.

ACKNOWLEDGMENTS
This research was supported by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER) funded through grant R305C120008 to the American Institutes for Research from the Institute for Education Sciences, U.S. Department of Education. Sorensen's contribution was supported in part by a pre-doctoral fellowship provided by the National Institute of Child Health and Human Development (T32-HD07376-25) through the Center for Developmental Science, University of North Carolina at Chapel Hill. The authors are grateful to Ibrahim Keita and Maria Laurito for research assistance.

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. doi:10.1086/508733.

Allensworth, Elaine M., and John Q. Easton. 2007. *What matters for staying on-track and graduating in Chicago Public Schools: A close look at course grades, failures, and attendance in the freshman year*. Chicago, IL: Consortium on Chicago School Research at the University of Chicago.

Balfanz, Robert, Lisa Herzog, and Douglas J. Mac Iver. 2007. Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist* 42(4):223–235. doi:10.1080/00461520701621079.

Berliner, David C. 2001. Learning about and learning from expert teachers. *International Journal of Educational Research* 35(5):463–482. doi:10.1016/S0883-0355(02)00004-6.

Berliner, David C. 2004. Teacher expertise. In *Teaching and learning in the secondary school*, edited by Bob Moon and Ann Shelton Mayes, pp. 107–113. New York: RoutledgeFalmer.

Betts, Julian R., Kim S. Rueben, and Anne Danenberg. 2000. *Equal resources, equal outcomes? The distribution of school resources and student achievement in California*. San Francisco, CA: Public Policy Institute of California.

Boyd, Donald, Pam Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2008. Who leaves? Teacher attrition and student achievement. NBER Working Paper No. 14022.

Byrnes, Vaughan, and Allen Ruby. 2007. Comparing achievement between K-8 and middle schools: A large-scale empirical study. *American Journal of Education* 114(1):101–135. doi:10.1086/520693.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2011. The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood. NBER Working Paper No. 17699.

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):778–820. doi:10.3368/jhr.XLI.4.778.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2007a. Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review* 26(6):673–682. doi:10.1016/j.econedurev.2007.10.002.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2007b. How and why do teacher credentials matter for student achievement? NBER Working Paper No. 12828.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2010. Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources* 45(3):655–681.

Clotfelter, Charles T., Helen F. Ladd, Jacob L. Vigdor, and Justin Wheeler. 2007. High poverty schools and the distribution of teachers and principals. *North Carolina Law Review* 85(5):1345–1379.

Cohen, David. 1996. Rewarding teachers for student performance. In *Rewards and reform: Creating educational incentives that work*, edited by Susan H. Fuhrman and Jennifer A. O'Day, pp. 60–114. San Francisco, CA: Josey-Bass.

Dee, Thomas S. 2004. Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics* 86(1):195–210. doi:10.1162/003465304323023750.

Duckworth, Angela L., and Martin E. P. Seligman. 2005. Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science* 16(12):939–944. doi:10.1111/j.1467-9280.2005.01641.x.

Goldhaber, Dan, Betheny Gross, and Daniel Player. 2011. Teacher career paths, teacher quality, and persistence in the classroom: Are public schools keeping their best? *Journal of Policy Analysis and Management* 30(1):57–87. doi:10.1002/pam.20549.

Hanushek, Eric A., John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin. 2005. The market for teacher quality. NBER Working Paper No. 11154.

Harris, Douglas N., and Tim R. Sass. 2011. Teacher training, teacher quality and student achievement. *Journal of Public Economics* 95(7–8):798–812. doi:10.1016/j.jpubeco.2010.11.009.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics* 24(3):411–482. doi:10.1086/504455.

Ingersoll, Richard, and Lisa Merrill. 2010. Who's teaching our children? *Educational Leadership* 67(8):14–20.

Isenberg, Eric, Bing-ru Teh, and Elias Walsh. 2015. Elementary school data issues for value-added models: Implications for research. *Journal of Research on Educational Effectiveness* 8(1):120–129. doi:10.1080/19345747.2014.974231.

Jackson, Kirabo C. 2012a. Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina. NBER Working Paper No. 18624.

Jackson, Kirabo C. 2012b. Teacher quality at the high-school level: The importance of accounting for tracks. NBER Working Paper No. 17722.

Kashdan, Todd B., Paul Rose, and Frank D. Fincham. 2004. Curiosity and exploration: Facilitating positive subjective experiences and personal growth opportunities. *Journal of Personality Assessment* 82(3):291–305. doi:10.1207/s15327752jpa8203_05.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review* 27(6):615–631. doi:10.1016/j.econedurev.2007.05.005.

Krieg, John M. 2006. Teacher quality and attrition. *Economics of Education Review* 25(1):13–27. doi:10.1016/j.econedurev.2004.09.004.

Labaree, David. 2000. On the nature of teaching and teacher education: Difficult practices that look easy. *Journal of Teacher Education* 51(3):228–233. doi:10.1177/002248710005100301l.

Ladd, Helen F. 2008. Teacher effects: What do we know? In *Teacher quality: Broadening and deepening the debate*, edited by Greg Duncan and James Spillane, pp. 3–26. Evanston, IL: Northwestern University.

Lankford, Hamilton, Susanna Loeb, and James Wyckoff. 2002. Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis* 24(1):37–62. doi:10.3102/01623737024001037.

Lleras, Christy. 2008. Do skills and behaviors in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings. *Social Science Research* 37(3):888–902. doi:10.1016/j.ssresearch.2008.03.004.

Marinell, William H., and Vanessa M. Coca, and the Middle School Teacher Turnover Project. 2013. *Who stays and who leaves? Findings from a three-part study of teacher turnover in NYC middle schools*. New York: The Research Alliance for New York City Schools.

Moffitt, Terrie E., Louise Arsenault, Daniel Belsky, Nigel Dickson, Robert J. Hancox, HonaLee Harrington, Renate Houts, Richie Poulton, Brent W. Roberts, Stephen Ross, Malcolm R. Sears, W. Murray Thomson, and Avshalom Caspi. 2011. A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences USA* 108(7):2693–2698.

North Carolina Department of Public Instruction (NCDPI). 2014. *Reliability of the North Carolina end-of-grade and end-of-course assessments*. Available www.ncpublicschools.org/docs/accountability/testing/eogeocreliabilities14.pdf. Accessed 22 March 2016.

Papay, John P., and Matthew A. Kraft. 2015. Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics* 130:105–119. doi:10.1016/j.jpubeco.2015.02.008.

Protik, Ali, Elias Walsh, Alexandra Resch, Eric Isenberg, and Emma Kopa. 2013. *Does tracking of students bias value-added estimates for teachers?* Mathematica Policy Research Working Paper No. 15.

Rivkin, Steven. 2006. Cumulative nature of learning and specification bias in education research. Unpublished paper, Amherst College.

Rockoff, Jonah E. 2004. The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review* 94(2):247–252. doi:10.1257/0002828041302244.

Steinberg, Laurence. 2005. Cognitive and affective development in adolescence. *Trends in Cognitive Sciences* 9(2):69–74. doi:10.1016/j.tics.2004.12.005.

Todd, Petra E., and Kenneth I. Wolpin. 2003. On the specification and estimation of production function for cognitive achievement. *Economic Journal* 113(485):F3–F33. doi:10.1111/1468-0297.00097.

Wiswall, Matthew. 2013. The dynamics of teacher quality. *Journal of Public Economics* 100:61–78. doi:10.1016/j.jpubeco.2013.01.006.

Xu, Zeyu, Umut Ozek, and Michael Hansen. 2014. Teacher performance trajectories in high and lower-poverty schools. *Educational Evaluation and Policy Analysis* 37(4):458–477. doi:10.3102/0162373714561513.

APPENDIX A: ADDITIONAL DATA

Table A.1. Math and Reading Scores

	Math			ELA		
Variables	**Model 1[a]**	**Model 2**	**2-Stage**	**Model 1[a]**	**Model 2**	**2-Stage**
Teacher experience						
No experience (base)						
Experience 1 year	0.0662**	0.0708**	0.0695**	0.0662**	0.0206**	0.0181**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)
Experience 2 years	0.0917**	0.0909**	0.0891**	0.0917**	0.0277**	0.0272**
	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)
Experience 3 years	0.1013**	0.0973**	0.0956**	0.1013**	0.0361**	0.0380**
	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)
Experience 4 years	0.1151**	0.1104**	0.1070**	0.1151**	0.0341**	0.0374**
	(0.005)	(0.007)	(0.006)	(0.005)	(0.007)	(0.007)
Experience 5 years	0.1372**	0.1290**	0.1280**	0.1372**	0.0395**	0.0454**
	(0.006)	(0.007)	(0.007)	(0.006)	(0.008)	(0.008)
Experience 6 years	0.1406**	0.1405**	0.1374**	0.1406**	0.0426**	0.0499**
	(0.006)	(0.008)	(0.008)	(0.006)	(0.009)	(0.008)
Experience 7 years	0.1530**	0.1513**	0.1492**	0.1530**	0.0494**	0.0567**
	(0.007)	(0.009)	(0.008)	(0.007)	(0.010)	(0.009)
Experience 8 years	0.1610**	0.1502**	0.1466**	0.1610**	0.0528**	0.0623**
	(0.007)	(0.010)	(0.009)	(0.007)	(0.011)	(0.010)
Experience 9 years	0.1567**	0.1471**	0.1431**	0.1567**	0.0596**	0.0686**
	(0.008)	(0.011)	(0.010)	(0.008)	(0.012)	(0.011)
Experience 10 years	0.1621**	0.1570**	0.1516**	0.1621**	0.0671**	0.0790**
	(0.009)	(0.011)	(0.010)	(0.009)	(0.013)	(0.011)
Experience 11 years	0.1688**	0.1694**	0.1637**	0.1688**	0.0722**	0.0856**
	(0.009)	(0.012)	(0.011)	(0.009)	(0.014)	(0.012)
Experience 12 years	0.1766**	0.1770**	0.1695**	0.1766**	0.0779**	0.0915**
	(0.010)	(0.013)	(0.012)	(0.010)	(0.015)	(0.013)
Experience 13–20 years	0.1716**	0.1752**	0.1679**	0.1716**	0.0765**	0.0920**
	(0.011)	(0.014)	(0.013)	(0.011)	(0.016)	(0.014)
Experience 21–27 years	0.1848**	0.1701**	0.1621**	0.1848**	0.0789**	0.1009**
	(0.012)	(0.016)	(0.014)	(0.012)	(0.018)	(0.016)
Experience 28-plus years	0.1711**	0.1545**	0.1468**	0.1711**	0.0790**	0.1042**
	(0.014)	(0.018)	(0.016)	(0.014)	(0.021)	(0.019)
Other time-varying teacher characteristics						
Lateral license	−0.0318**	−0.0358**	−0.0322**	−0.0108	−0.0208***	−0.0260*
	(0.008)	(0.010)	(0.009)	(0.009)	(0.011)	(0.011)
Same race as student	0.0049**	0.0068**	0.0071**	−0.0033*	0.0007	−0.0006
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Same gender as student	0.0123**	0.0068**	0.0072**	0.0015	0.0013	0.0101**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Student characteristics						
Male	−0.0102**			−0.0020		
	(0.001)			(0.001)		
Black	−0.0320**			−0.0681**		
	(0.002)			(0.002)		
Hispanic	0.0301**			0.0075**		
	(0.002)			(0.002)		
Other race	0.0393**			0.0029		
	(0.002)			(0.002)		
Age in grade 3	−0.0499**			−0.0437**		
	(0.001)			(0.001)		
Parent college graduate (base)						
Parent community college graduate	−0.0252**			−0.0201**		
	(0.001)			(0.002)		
Parent high school graduate	−0.0411**			−0.0436**		
	(0.001)			(0.001)		
Parent high school drop out	−0.0618**			−0.0665**		
	(0.003)			(0.003)		

Table A.1. Continued.

	Math			ELA		
Variables	**Model 1[a]**	**Model 2**	**2-Stage**	**Model 1[a]**	**Model 2**	**2-Stage**
Limited English	−0.0049* (0.002)			−0.0613** (0.003)		
Gifted in Math	0.1087** (0.002)			0.0672** (0.002)		
Gifted in Reading	0.0779** (0.002)			0.0883** (0.002)		
Special needs	−0.0486** (0.002)			−0.0564** (0.002)		
Subsidized lunch	−0.0437** (0.001)			−0.0324** (0.001)		
Repeated grade	−0.0748** (0.003)			−0.0356** (0.003)		
School change	−0.0319** (0.002)	−0.0168** (0.002)	−0.0102** (0.002)	−0.0229** (0.002)	−0.0033 (0.002)	−0.0006 (0.002)
Structural school change	−0.0976** (0.003)	−0.0088** (0.003)	0.0135** (0.002)	−0.0858** (0.003)	−0.0016 (0.004)	0.0101** (0.002)
Lagged test score	0.7611** (0.001)			0.8024** (0.001)		
Classroom characteristics						
Class size between 21 and 30 (base)						
Class size <=5	−0.0104* (0.005)	−0.0193** (0.007)	−0.0018 (0.006)	−0.0336** (0.005)	−0.0089 (0.007)	−0.0106 (0.007)
Class size between 6 and 10	0.0186** (0.003)	0.0051 (0.004)	0.0134** (0.004)	−0.0124** (0.003)	−0.0006 (0.004)	−0.0016 (0.004)
Class size between 11 and 20	0.0093** (0.001)	0.0075** (0.001)	0.0117** (0.001)	−0.0043** (0.001)	0.0006 (0.002)	0.0007 (0.002)
Class size between 31 and 40	0.0101** (0.002)	−0.0019 (0.003)	−0.0050*** (0.003)	0.0045*** (0.002)	−0.0118** (0.003)	−0.0109** (0.003)
Class size $\geq$ 41	−0.0260** (0.006)	−0.0202** (0.007)	−0.0254** (0.007)	−0.0025 (0.005)	−0.0014 (0.006)	−0.0010 (0.005)
Percent nonwhite	−0.0164** (0.004)	−0.0198** (0.006)	−0.0045 (0.006)	−0.0157** (0.005)	−0.0035 (0.006)	−0.0021 (0.006)
Percent eligible for subsidized lunch	−0.1420** (0.004)	−0.0370** (0.006)	0.0057 (0.006)	−0.1071** (0.005)	−0.0377** (0.006)	−0.0391** (0.006)
Percent parent college grad (base)						
Percent parent comm. college graduate	−0.0059 (0.004)	−0.0172** (0.007)	0.0072 (0.006)	−0.0022 (0.005)	−0.0086 (0.007)	−0.0093 (0.007)
Percent parent high school graduate	0.0081*** (0.004)	−0.0326** (0.007)	−0.0001 (0.006)	−0.0001 (0.004)	−0.0265** (0.007)	−0.0283** (0.007)
Percent parent high school dropout	−0.0230* (0.012)	−0.0335* (0.016)	0.0019 (0.015)	−0.0573** (0.012)	−0.0725** (0.016)	−0.0704** (0.016)
Average lagged test score	0.0054** (0.001)	−0.0756** (0.002)	−0.0374** (0.002)	0.0260** (0.002)	−0.0712** (0.002)	−0.0712** (0.002)
Parent education imputed	0.0116** (0.002)			0.0014** (0.002)		
Year × Grade FE	YES	YES	YES	YES	YES	YES
School FE	YES	NO	NO	YES	NO	NO
Student FE	NO	YES	YES	NO	YES	YES
Teacher FE	YES	YES	YES	YES	YES	YES
Observations	1,237,088	1,237,088	1,322,296	1,241,452	1,241,452	1,316,972
R^2	0.712	0.936	0.947	0.689	0.925	0.951

Notes: Robust standard errors in parentheses.

[a]Model 1 estimates come from errors-in-variables regression which corrects for measurement error in the lagged test score term.

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.1$.

Table A.2. Test of Differential Returns to Experience by Teacher Cohort (Model 2)

Teacher Experience	Math	ELA
1 year	0.0694**	0.0172**
	(0.004)	(0.004)
2 years	0.0880**	0.0244**
	(0.005)	(0.005)
3 years	0.0939**	0.0339**
	(0.006)	(0.006)
4 years	0.1069**	0.0313**
	(0.006)	(0.007)
5 years	0.1233**	0.0381**
	(0.009)	(0.009)
× 2009–2011 cohorts	0.0061	−0.0007
	(0.006)	(0.007)
6 years	0.1372**	0.0409**
	(0.008)	(0.009)
7 years	0.1439**	0.0555**
	(0.010)	(0.011)
× 2009–2011 cohorts	0.0062	−0.0128+
	(0.007)	(0.007)
8 years	0.1446**	0.0517**
	(0.010)	(0.011)
9 years	0.1382**	0.0489**
	(0.012)	(0.013)
× 2009–2011 cohorts	0.0033	0.0106
	(0.006)	(0.007)
10 years	0.1486**	0.0648**
	(0.011)	(0.013)
11 years	0.1555**	0.0604**
	(0.013)	(0.015)
× 2009–2011 cohorts	0.0066	0.0111
	(0.007)	(0.008)
12 years	0.1648**	0.0709**
	(0.013)	(0.015)
13–20 years	0.1628**	0.0686**
	(0.014)	(0.016)
21–27 years	0.1578**	0.0742**
	(0.016)	(0.018)
28-plus years	0.1455**	0.0761**
	(0.018)	(0.020)
Teacher FE	YES	YES
Student FE	YES	YES
Grade × Year FE	YES	YES
Observations	1,317,102	1,322,296
R^2	0.936	0.925

Notes: The table tests for the possibility of different returns to experience by adjacent cohorts (those with a certain experience level in 2007–08 versus 2009–11).

$^{**}p < 0.01$; $^{***}p < 0.1$.

Table A.3. Non-test Student Outcomes: ELA Sample (Model 2)

Variables	Absences	Classroom Offenses	Homework Effort	Free Reading
Teacher credentials				
No experience (base)				
Experience 1 year	−0.0200** (0.004)	−0.0059* (0.003)	0.0017 (0.024)	0.0113* (0.005)
Experience 2 years	−0.0448** (0.005)	−0.0054 (0.004)	0.0243 (0.028)	0.0181** (0.006)
Experience 3 years	−0.0298** (0.006)	0.0115** (0.004)	0.0343 (0.033)	0.0260** (0.007)
Experience 4 years	−0.0344** (0.007)	0.0139** (0.005)	−0.0081 (0.038)	0.0238** (0.008)
Experience 5 years	−0.0410** (0.008)	0.0062 (0.005)	0.0145 (0.042)	0.0256** (0.009)
Experience 6 years	−0.0386** (0.008)	0.0037 (0.006)	−0.0063 (0.047)	0.0221* (0.010)
Experience 7 years	−0.0622** (0.009)	0.0068 (0.007)	0.0543 (0.051)	0.0337** (0.011)
Experience 8 years	−0.0648** (0.010)	0.0021 (0.007)	0.0345 (0.056)	0.0379** (0.012)
Experience 9 years	−0.0566** (0.011)	0.0024 (0.008)	0.0494 (0.060)	0.0427** (0.013)
Experience 10 years	−0.0591** (0.012)	0.0060 (0.008)	0.0876 (0.065)	0.0448** (0.014)
Experience 11 years	−0.0580** (0.013)	0.0034 (0.009)	0.0557 (0.069)	0.0325* (0.016)
Experience 12 years	−0.0624** (0.013)	0.0006 (0.010)	0.0916 (0.074)	0.0456** (0.017)
Experience 13–20 years	−0.0815** (0.015)	−0.0023 (0.010)	0.1202 (0.080)	0.0392* (0.018)
Experience 21–27 years	−0.1450** (0.017)	−0.0106 (0.012)	0.1638+ (0.092)	0.0497* (0.021)
Experience 28-plus years	−0.1408** (0.019)	−0.0153 (0.014)	0.2270* (0.104)	0.0425*** (0.023)
Regular license (base)				
Lateral license	0.1030** (0.010)	0.0272** (0.007)	0.0054 (0.058)	−0.0283* (0.013)
Student characteristics				
School change	−0.0170** (0.002)	0.0074** (0.001)	−0.0220+ (0.013)	−0.0007 (0.003)
Structural school change	−0.0260** (0.003)	0.0024 (0.002)	0.0283 (0.020)	−0.0049 (0.004)
Classroom characteristics				
Class size from 21 to 30 (base)				
Class size ≤ 5	−0.0316** (0.006)	0.0031 (0.004)	−0.0830 (0.062)	0.0011 (0.009)
Class size between 6 and 10	−0.0256** (0.004)	0.0069* (0.003)	0.0186 (0.031)	0.0088*** (0.005)
Class size between 11 and 20	−0.0017 (0.001)	−0.0003 (0.001)	−0.0047 (0.008)	−0.0008 (0.002)
Class size between 31 and 40	−0.1065** (0.005)	−0.0041* (0.002)	0.0138 (0.014)	−0.0030 (0.003)
Class size ≥ 41	−0.0268** (0.006)	−0.0095* (0.004)	0.0258 (0.033)	−0.0057 (0.006)
% Nonwhite	−0.0128* (0.006)	−0.0019 (0.004)	0.0798* (0.037)	0.0052 (0.007)
% Eligible subsidized lunch	−0.0316** (0.006)	0.0104* (0.004)	−0.2669** (0.037)	0.0077 (0.007)

Table A.3. Continued.

Variables	Absences	Classroom Offenses	Homework Effort	Free Reading
% Parent college grad (base)				
% Parent comm. college grad	0.0377** (0.006)	0.0058 (0.005)	−0.1803** (0.042)	−0.0030 (0.008)
% Parent high school grad	0.0165** (0.006)	−0.0035 (0.004)	−0.2200** (0.042)	−0.0160*** (0.008)
% Parent high school dropout	0.0250*** (0.015)	0.0042 (0.011)	−0.1559 (0.108)	−0.0363*** (0.021)
Mean peer lagged score	−0.0168** (0.002)	−0.0157** (0.001)	0.0064 (0.010)	0.0056* (0.002)
Grade × Year FE	YES	YES	YES	YES
School FE	NO	NO	NO	NO
Student FE	YES	YES	YES	YES
Teacher FE	YES	YES	YES	YES
Observations	1,329,206	1,332,977	976,278	980,163
R^2	0.721	0.731	0.730	0.804

Notes: Robust standard errors in parentheses.

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.1$.

Table A.4. Non-test Student Outcomes: Math Sample (Model 2)

Variables	Absences	Classroom Offenses	Homework Effort
Teacher credentials			
No experience (base)			
Experience 1 year	−0.0194** (0.004)	−0.0168** (0.003)	0.0064 (0.025)
Experience 2 years	−0.0381** (0.005)	−0.0192** (0.004)	0.0364 (0.029)
Experience 3 years	−0.0740** (0.006)	−0.0130** (0.004)	0.0442 (0.033)
Experience 4 years	−0.0222** (0.007)	−0.0190** (0.005)	0.0486 (0.037)
Experience 5 years	−0.0587** (0.007)	−0.0152** (0.005)	0.0434 (0.042)
Experience 6 years	−0.0319** (0.008)	−0.0213** (0.006)	0.0269 (0.046)
Experience 7 years	−0.0561** (0.009)	−0.0224** (0.006)	0.0380 (0.051)
Experience 8 years	−0.0805** (0.010)	−0.0188** (0.007)	0.0345 (0.055)
Experience 9 years	−0.0625** (0.011)	−0.0230** (0.007)	0.0549 (0.059)
Experience 10 years	−0.0474** (0.011)	−0.0184* (0.008)	0.0255 (0.063)
Experience 11 years	−0.0557** (0.012)	−0.0180* (0.009)	0.0639 (0.068)
Experience 12 years	−0.1114** (0.013)	−0.0157*** (0.009)	0.0805 (0.072)
Experience 13–20 years	−0.0924** (0.014)	−0.0204* (0.010)	0.0742 (0.077)
Experience 21–27 years	−0.1070** (0.016)	−0.0233* (0.011)	0.0431 (0.087)
Experience 28-plus years	−0.1154** (0.018)	−0.0181 (0.013)	−0.0035 (0.099)

Table A.4. Continued.

Variables	Absences	Classroom Offenses	Homework Effort
Regular license (base)			
Lateral license	0.0053	−0.0216	0.0754
	(0.010)	(0.026)	(0.055)
Student characteristics			
School change	−0.0203**	0.0299**	−0.0092
	(0.002)	(0.005)	(0.013)
Structural school change	−0.0266**	0.0261**	0.0153
	(0.003)	(0.007)	(0.020)
Classroom characteristics			
Class size between 21 and 30 (base)			
Class size ≤ 5	−0.0350**	0.0978**	−0.0739
	(0.006)	(0.023)	(0.064)
Class size between 6 and 10	−0.0240**	0.0447**	0.0145
	(0.004)	(0.013)	(0.032)
Class size between 11 and 20	−0.0111**	0.0048	0.0091
	(0.001)	(0.004)	(0.008)
Class size between 31 and 40	0.0258**	−0.0265**	0.0301*
	(0.007)	(0.007)	(0.014)
Class size ≥ 41	−0.1833**	−0.0228	0.0557***
	(0.007)	(0.017)	(0.033)
% Nonwhite	0.0035	−0.0716**	0.1167**
	(0.006)	(0.017)	(0.037)
% Eligible subsidized lunch	−0.0119*	0.1085**	−0.2179**
	(0.006)	(0.017)	(0.038)
% Parent college grad (base)			
% Parent community college grad	0.0153*	0.0213	−0.1341**
	(0.007)	(0.019)	(0.042)
% Parent high school grad	0.0227**	−0.0420*	−0.1247**
	(0.007)	(0.019)	(0.042)
% Parent high school dropout	0.0132	0.0019	−0.0411
	(0.016)	(0.045)	(0.109)
Mean peer lagged score	−0.0003	−0.0753**	0.1024**
	(0.002)	(0.004)	(0.009)
Grade × Year FE	YES	YES	YES
School FE	NO	NO	NO
Student FE	YES	YES	YES
Teacher FE	YES	YES	YES
Observations	1,319,672	1,325,142	969,364
R^2	0.716	0.733	0.731

Notes: Robust standard errors in parentheses.

*p < 0.05; **p < 0.01; ***p < 0.1.

Table A.5. Returns to Experience for Absence Indicators with Alternative Thresholds (Model 2)

	Absences > 3 (25th percentile)		Absences > 6 (Median)		Absences > 17 (90th percentile)	
Teacher Experience	**ELA**	**Math**	**ELA**	**Math**	**ELA**	**Math**
No experience (base)						
1 year	−0.0030	−0.0152**	−0.0142**	−0.0214**	−0.0258**	−0.0316**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
2 years	−0.0164**	−0.0236**	−0.0343**	−0.0326**	−0.0513**	−0.0424**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)

Table A.5. Continued.

	Absences > 3 (25th percentile)		Absences > 6 (Median)		Absences > 17 (90th percentile)	
Teacher Experience	**ELA**	**Math**	**ELA**	**Math**	**ELA**	**Math**
3 years	−0.0054	−0.0357**	−0.0224**	−0.0623**	−0.0368**	−0.0900**
	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)
4 years	0.0005	−0.0164**	−0.0215**	−0.0247**	−0.0469**	−0.0271**
	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)
5 years	−0.0112	−0.0361**	−0.0342**	−0.0545**	−0.0499**	−0.0697**
	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)
6 years	−0.0094	−0.0226**	−0.0301**	−0.0286**	−0.0490**	−0.0429**
	(0.008)	(0.008)	(0.009)	(0.008)	(0.007)	(0.007)
7 years	−0.0197*	−0.0304**	−0.0545**	−0.0476**	−0.0749**	−0.0711**
	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
8 years	−0.0178+	−0.0351**	−0.0502**	−0.0696**	−0.0731**	−0.0992**
	(0.010)	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)
9 years	−0.0203*	−0.0338**	−0.0464**	−0.0608**	−0.0675**	−0.0764**
	(0.010)	(0.010)	(0.011)	(0.011)	(0.010)	(0.009)
10 years	−0.0107	−0.0207***	−0.0439**	−0.0504**	−0.0694**	−0.0645**
	(0.011)	(0.011)	(0.012)	(0.012)	(0.010)	(0.010)
11 years	−0.0143	−0.0216***	−0.0419**	−0.0531**	−0.0770**	−0.0652**
	(0.012)	(0.012)	(0.013)	(0.013)	(0.011)	(0.011)
12 years	−0.0193	−0.0523**	−0.0416**	−0.0973**	−0.0906**	−0.1291**
	(0.013)	(0.012)	(0.014)	(0.013)	(0.012)	(0.012)
13–20 years	−0.0311*	−0.0409**	−0.0578**	−0.0783**	−0.1077**	−0.1088**
	(0.014)	(0.013)	(0.015)	(0.014)	(0.013)	(0.013)
21–27 years	−0.0564**	−0.0440**	−0.1069**	−0.0796**	−0.1880**	−0.1220**
	(0.016)	(0.015)	(0.017)	(0.016)	(0.015)	(0.014)
28-plus years	−0.0467**	−0.0522**	−0.0944**	−0.0916**	−0.1843**	−0.1350**
	(0.018)	(0.017)	(0.020)	(0.019)	(0.017)	(0.016)
Observations	1,329,323	1,319,725	1,329,323	1,319,725	1,329,323	1,319,725
R^2	0.698	0.698	0.718	0.716	0.713	0.709

Notes: Standard errors in parentheses.

$^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.1$.

Table A.6. Returns to Experience for Offense Indicators with Alternative Thresholds (Model 2)

	Offenses >1		Offenses >2	
Experience	**ELA**	**Math**	**ELA**	**Math**
No experience (base)				
1 year	−0.0064**	−0.0133**	−0.0055**	−0.0104**
	(0.002)	(0.002)	(0.002)	(0.002)
2 years	−0.0053*	−0.0123**	−0.0057**	−0.0107**
	(0.003)	(0.003)	(0.002)	(0.002)
3 years	0.0045	−0.0072*	0.0021	−0.0066**
	(0.003)	(0.003)	(0.003)	(0.002)
4 years	0.0042	−0.0127**	0.0005	−0.0096**
	(0.004)	(0.003)	(0.003)	(0.003)
5 years	0.0026	−0.0079*	−0.0001	−0.0051
	(0.004)	(0.004)	(0.003)	(0.003)
6 years	−0.0034	−0.0107*	−0.0050	−0.0082*
	(0.004)	(0.004)	(0.004)	(0.004)

Table A.6. Continued.

	Offenses >1		Offenses >2	
Experience	ELA	Math	ELA	Math
7 years	−0.0036 (0.005)	−0.0139** (0.005)	−0.0049 (0.004)	−0.0093* (0.004)
8 years	−0.0049 (0.005)	−0.0088*** (0.005)	−0.0076*** (0.004)	−0.0043 (0.004)
9 years	−0.0063 (0.006)	−0.0138* (0.006)	−0.0068 (0.005)	−0.0094* (0.005)
10 years	−0.0049 (0.006)	−0.0083 (0.006)	−0.0073 (0.005)	−0.0055 (0.005)
11 years	−0.0058 (0.007)	−0.0090 (0.006)	−0.0081 (0.005)	−0.0058 (0.005)
12 years	−0.0077 (0.007)	−0.0074 (0.007)	−0.0085 (0.006)	−0.0051 (0.006)
13–20 years	−0.0104 (0.008)	−0.0088 (0.007)	−0.0098 (0.006)	−0.0055 (0.006)
21–27 years	−0.0182* (0.009)	−0.0054 (0.008)	−0.0130+ (0.007)	−0.0058 (0.007)
28-plus years	−0.0169*** (0.010)	−0.0003 (0.009)	−0.0133 (0.008)	−0.0013 (0.008)
Observations	1,329,323	1,319,725	1,329,323	1,319,725
R^2	0.696	0.699	0.681	0.684

Notes: Standard errors in parentheses.

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.1$.