Educational Evaluation and Policy Analysis
December 2015, Vol. 37, No. 4, pp. 458–477
DOI: 10.3102/0162373714561513

Teacher Performance Trajectories in High- and Lower-Poverty Schools

Zeyu Xu
Umut Özek
Michael Hansen
American Institutes for Research/CALDER

This study explores whether teacher performance trajectory over time differs by school-poverty settings. Focusing on elementary school mathematics teachers in North Carolina and Florida, we find no systematic relationship between school student poverty rates and teacher performance trajectories. In both high- (≥60% free/reduced-price lunch [FRPL]) and lower-poverty (<60% FRPL) schools, teacher performance improves the fastest in the first 5 years and then flattens out in years 5 to 10. Teacher performance growth resumes between year 10 and 15 in North Carolina but remains flat in Florida. In both school-poverty settings, there is a significant variation in teacher performance trajectories. Among novice and early-career teachers, the fastest-growing teachers (75th percentile) improve by 0.04 standard deviations more in student gain scores annually than slower teachers (25th percentile). In both school settings, novice teachers who started with low effectiveness also grew at a slower rate in the next 5 years than novice teachers with higher initial effectiveness. Our findings suggest that the lack of productivity "return" to experience in high-poverty schools reported in the literature is unlikely to be the result of differential teacher learning in high- and lower-poverty schools.

Keywords: *economics of education, educational policy, educational reform, evaluation, school/teacher effectiveness*

Evaluating the productivity of schoolteachers has become a focal point in recent policy efforts to improve the nation's public school system. Motivated in part by the U.S. Department of Education's recent Race to the Top Initiative, 36 states have updated their teacher evaluation policies since 2009 to take a more active role in managing workforce quality (National Council on Teacher Quality, 2012). In light of this growing public interest, unanswered questions about the contexts in which teachers develop their human capital have come to the fore.

One of these unanswered questions is whether teachers' experience–productivity profiles systematically differ across school settings. This issue is significant because the interaction between teachers' productivity growth and the school setting where teachers initially gain their experience has the potential to either inadvertently reinforce or compensate for preexisting inequalities when high stakes are attached to performance. Sass, Hannaway, Xu, Figlio, and Feng's (2012) examination of the distributions of teacher quality in high- and low-poverty schools in North Carolina and Florida motivates the investigation presented here. Sass et al. (2012) presents cross-sectional evidence showing no association between experience and value-added productivity in high-poverty schools in both states. However, their study was unable to address whether this finding was due to the composition of the teacher workforce in those schools (due to non-random selection into and out of the schools) or whether it was due to systemic differences in the development of teachers' productivity in high-poverty environments.

Therefore, we investigate this surprising finding further by asking the following research question: In comparison with teachers in lower-poverty schools, do teachers in high-poverty schools exhibit different productivity trajectories (i.e., changes in performance levels over time) during various stages of their careers? To address this question, we focus on elementary school mathematics teachers and utilize longitudinal administrative data from North Carolina and Florida that span 11 and 8 years, respectively. We first estimate the annual value-added for each teacher and then estimate teacher-specific performance trajectories in the second stage using a three-level hierarchical linear model (HLM) that takes into account changes in classroom and school characteristics over time as well as sampling errors associated with value-added estimates. We address our research question by comparing the mean and variation in estimated teacher trajectories between high- and lower-poverty schools.

In summary of our results, we find no systematic relationship between school student poverty rates and teacher performance trajectories. In both high- (schools with 60% or more of their students eligible for free/reduced-price lunch [FRPL]) and lower-poverty (schools with less than 60% FRPL-eligible students) schools, teacher performance improves the fastest in the first 5 years and then flattens out in years 5 to 10. Teacher performance growth resumes between year 10 and 15 in North Carolina but remains flat in Florida.

In addition, in high- and lower-poverty schools alike, there is a significant variation in teachers' performance trajectories at all career stages. Among novice and early-career teachers, the fastest-growing teachers (75th percentile) improve more than slower teachers (25th percentile) by about 0.04 standard deviations in student gain scores annually, roughly equivalent to half a year of additional growth during an average teacher's first 3 to 5 years of teaching. In both school settings, novice teachers who start with low effectiveness also grow at a slower rate than novice teachers with higher initial effectiveness.

Our finding that the growth trajectory of teacher productivity does not differ by school-poverty setting remains unchanged when we use alternative FRPL cutoff values to define "high-poverty" status (≥60% vs. <40%, or ≥70% vs. <30%). Our study suggests that the lack of a "return" to experience in high-poverty schools reported in Sass et al. (2012) is unlikely to be the result of differential teacher learning in high- and lower-poverty schools; rather, it is more likely the result of non-random sorting among less-effective experienced teachers into the workforce of high-poverty schools.

The rest of this article is organized into six sections. In the following section, we situate our study in the current research background. Sections "Data and Samples" and "Method" present the data and methods we use. Section "Findings" presents our results and Section "Summary and Discussion" concludes our study.

Research Background

Research in recent years about the influence of teacher productivity on student learning has produced three key findings, which have emerged as a consensus. First, the variability of teacher productivity across the workforce is large, with the differential effect of having an effective teacher versus an ineffective teacher greater than the effect sizes associated with other educational interventions, such as class size reduction (for recent reviews of the findings in this literature, see Hanushek & Rivkin, 2010; Staiger & Rockoff, 2010). Second, differences in past teacher productivity are associated with not only short-term student cognitive gains but also long-term student outcomes, including college enrollment, future wages, and other non-cognitive outcomes (Chetty, Friedman, & Rockoff, 2014). And third, teacher characteristics most commonly observed in administrative educational data (e.g., teachers' credentials and experience) are only weakly associated with differences in teacher value-added productivity (Aaronson, Barrow, & Sander, 2007; Koedel & Betts, 2007).

Consequently, research in recent years has moved toward investigating how workforce management policies may be crafted to identify and retain the best teachers while removing ineffective teachers from the classroom (e.g., Goldhaber & Hansen, 2010; Gordon, Kane, & Staiger, 2006; Staiger & Rockoff, 2010). Yet, proposed approaches to manage workforce quality generally take teacher productivity as a given and do not address the context in which it develops over time. This is a consequential omission, given the difficulties disadvantaged schools have in retaining their best teachers (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2008; Hanushek, Kane, & Rivkin, 2004; Scafidi, Sjoquist, & Stinebrickner, 2007). If the development of teacher productivity systematically differed in disadvantaged schools compared with non-disadvantaged schools, the adoption of policies that make high-stakes decisions about teachers' careers based on performance has particular implications for disadvantaged schools: On one hand, if teachers, on average, increase their productivity more quickly when staffed in disadvantaged schools than they are expected in other schools, then the adoption of such policies could indirectly attract teachers to such schools and thereby reduce inequalities. Conversely, if teachers do not improve their practice as quickly in disadvantaged schools, then high-stakes workforce policies may inadvertently reinforce inequalities by deterring teachers from such schools even more.

Prior studies on the returns to teachers' experience have not directly addressed this issue. To begin, literature shows that teachers grow the most in the first few years of their careers (Rockoff, 2004), and more recent evidence suggests that returns to experience, on average, may continue to be significantly positive throughout their careers (Papay & Kraft, 2011; Wiswall, 2010). Using North Carolina data, Kraft and Papay's (in press) recent analysis suggests that the school context may be an important factor in developing teachers' productivity. However, their explanatory variables of interest in these authors' analysis are from a Working Conditions Survey that captures the level of administrative and colleague support rather than differences in schools' student socioeconomic composition. Moreover, studies on changes in teacher productivity over time have investigated how productivity *levels* vary across different school settings but have not investigated systematically different *trajectories* by settings (e.g., Jackson, 2010; Jackson & Bruegmann, 2009; Xu, Ozek, & Corritore, 2012). Hence, this topic represents an important gap in the research base.

Recent findings from Sass et al. (2012), in particular, show an empirical puzzle, which motivates our inquiry here. Using two states' data, the authors find that the distributions of teacher performance in low-poverty schools are, surprisingly, virtually identical to those in high-poverty schools. The notable exception, however, is at the lower end of the teacher performance distributions. The authors report that the least effective teachers in high-poverty schools are outperformed by the least effective teachers in low-poverty schools. They investigate this phenomenon further and, based on cross-sectional results, find no significant relationship between teacher performance and experience in high-poverty schools, which diverged from that in lower-poverty schools. The authors hypothesize that this empirical result is likely due to one of two possible sources: Either human capital development takes on a significantly different trajectory for teachers in high-poverty schools or low-quality experienced teachers are systematically sorted into high-poverty schools (via a "dance of the lemons" mechanism; Miller & Chait, 2008).

These two competing hypotheses invite inquiry into whether trajectories of teachers' performance (i.e., the slopes in their productivity–experience profiles) may systematically differ in high-poverty schools, as has already been shown to be the case with the average level of teacher performance in such schools. These different slopes could feasibly arise from the outset of a teacher's career. For example, teachers in high-poverty schools may have lower returns to experience in the initial years to begin with and never catch up to their counterparts in low-poverty schools. Alternatively, after having similar rate of growth for several years as new teachers, teachers in high-poverty schools may be more likely to plateau in their performance or even decline (i.e., "burnout"), whereas teacher performance in low-poverty schools may be more likely to continually improve over time.

Data and Samples

This study draws on longitudinal student and teacher data from North Carolina (1998–1999 through 2008–2009) and Florida (2002–2003 through 2009–2010). In both states, we focus on fourth- and fifth-grade mathematics teachers. The relatively long study periods (11 and 8 years) increase our chance of observing the same teacher repeatedly over a longer period of time and allow us to estimate their growth more reliably. In addition, using large-scale data from two distinct public education systems would greatly strengthen the findings of our study if consistent patterns were to emerge despite all the differences in student testing, curriculum, practices, and policy contexts in general. Analytic samples were drawn from the populations in two steps: First, we extracted student–teacher-level samples to estimate teacher annual performance using value-added models. Next, we constructed teacher-year samples to investigate teacher performance growth over time.

Samples Used to Estimate Teacher Performance

In both states, end-of-grade (EOG) mathematics tests are administered annually to elementary school students starting from the third grade. This allows us to estimate value-added for teachers in Grades 4 and 5, using previous year's student test scores to control for student prior performance.

We restricted our North Carolina samples to teachers and students in "self-contained" classrooms, mainly because North Carolina did not have direct instructor–student links prior to 2006–2007. For those earlier years, therefore, we need to verify whether test proctors are actual instructors. Specifically, information on students and teachers is contained in two separate files in North Carolina. The "instructional classes" file is a *classroom-level* file that includes aggregate student characteristics and instructor IDs. The "test score" file is a *student-level* file that includes test proctor IDs as well as student test scores and student characteristics. As a result, instructors are not linked directly to individual students; only proctors are. We can verify whether proctors are indeed instructional teachers by comparing aggregate student characteristics (percent male, percent White, and class size) in the instructional classrooms and those in the test classrooms. To do this, we (a) aggregated individuals in the test score file into test classrooms and (b) linked the test and instructional classes (now both are classroom-level files) by local education agency (LEA; district), school ID, and teacher ID. If the two sets of student characteristics were sufficiently similar (defined as the mean squared difference of the three classroom characteristics), we would confirm a test proctor to be the actual instructor. This verification process has been successfully implemented in a number of earlier studies using the same North Carolina data (see Goldhaber & Anthony, 2007; Clotfelter, Ladd, & Vigdor, 2010; Xu, Hannaway, & Taylor, 2011, for more details).

The Florida data, however, contain student–teacher links throughout the study period, and so we did not need to restrict our samples to those in self-contained classrooms. However, to attribute student learning gains to teachers more accurately, we restricted our samples to teachers and students in "core" mathematics courses, which are defined as those that more than 50% of students in a given grade take in a given school. We further excluded students "exposed" to more than one teacher in a given subject during a school year.

After these steps, we were able to identify about 42,000 unique elementary school mathematics teachers in North Carolina, among which 32,000 teachers could be verified as classroom instructors. In Florida, we identified about 36,000 unique elementary school mathematics teachers. To reduce potential sample heterogeneity, we further restricted our samples by (a) removing charter school teachers, (b) removing students and teachers who changed schools during a school year (about 2%–4% of observations), (c) keeping classrooms (in the analytic sample) with 10 to 40 students, and (d) removing classrooms with more than 50% special education students. The final samples used to estimate teacher value-added include around 21,000 and 30,000 elementary school mathematics teachers in North Carolina and Florida, respectively (Table 1).

Samples Used to Estimate Teacher Growth

To explore how teacher performance improves with experience and how the rate of improvement may vary by school student poverty rates, we divided teachers with valid value-added estimates into three groups according to their

TABLE 1

Number of Elementary School Mathematics Teachers in State and Subsamples, by Sample Restriction Steps

	North Carolina	Florida
Samples used to estimate value-added		
Teachers of relevant classes	41,691	36,446
Teachers linked to students	32,205	36,446
Eliminate charter school classes	22,254	34,717
Keep classes with 10–40 students who has no missing values on student and teacher variables	21,119	29,989
Samples used to estimate teacher growth		
Novice teachers (cohorts with 0–1 year of experience)	5,883	8,295
Stayed in the same school-poverty setting[a]	3,317	7,469
Early-career teachers (cohorts with 5–6 years of experience)	4,117	5,068
Stayed in the same school-poverty setting[a]	2,536	4,490
Mid-career teachers (cohorts with 10–11 years of experience)	2,775	2,676
Stayed in the same school-poverty setting[a]	1,846	2,441

[a]Schools with 60% or more students eligible for free/reduced-price lunch are defined as high-poverty schools. Otherwise, they are defined as lower-poverty schools.

experience levels. Research consistently finds disproportionate representation of inexperienced teachers in high-poverty schools and that returns to experience are the most evident during the earlier years. Therefore, a direct comparison of the average growth rates among *all* teachers in high- and lower-poverty schools would be misleading. By stratifying teachers according to their experience levels and examining subgroups of teachers at comparable career stages, we can remove this potential confounding factor.

The "novice teacher" group includes teachers with 0 or 1 year of experience who can be followed for up to 5 years in our data. Although we estimate performance trajectories for all teachers who have sufficient data, we are particularly interested in those who have taught continuously in a high-poverty school setting (≥60% of students eligible for FRPL) or in a lower-poverty school setting (<60% of students eligible for FRPL) during this 5-year period. The "early-career teacher" group includes teachers with 5 or 6 years of experience who can be followed for up to 5 years after that. Similarly, we focus on those who have stayed in the same school-poverty setting during this period. Finally, the "mid-career teacher" group includes teachers with 10 or 11 years of experience who can be followed for up to 5 years. We again focus on those who have not switched school-poverty settings. These three groups roughly correspond to key stages in a teacher's career that have been documented to have distinct performance trajectories. The sample sizes of each group are reported in the bottom panel of Table 1.

Method

We conduct our analyses in three stages. In the first stage, we estimate the annual performance of each individual teacher within the "value-added" framework. The resulting teacher value-added estimates are then used as the dependent variable in Stage 2, where we estimate teacher-specific performance trajectories over time using a three-level HLM. In the final stage of our analysis, we focus on a subsample of teachers who have not switched school-poverty settings during the study period and compare their performance trajectories over time. We discuss each analytic stage in detail below.

Evaluating Teacher Performance

We first estimate teacher annual performance using "value-added" scores. In a value-added framework, education is viewed as a cumulative process. It is assumed that time-lagged student achievement sufficiently captures all historical inputs and heritable endowments in the education process (Todd & Wolpin, 2003), thus separating the current teacher's contribution to student learning from the effects of teachers and other education inputs in earlier years. To account for other contributing factors to student learning in the current period and to mitigate non-random sorting among teachers, students, and schools, value-added models typically also control for a number of observable student characteristics in addition to lagged student test scores. We estimate teacher value-added using the following teacher fixed-effects model[1]:

$$A_{it} - A_{it-1} = \mathbf{T}_{it}\beta_{it} + X_{it}\gamma + \varepsilon_{it}. \tag{1}$$

$A_{it} - A_{it-1}$ is the student test score gain between year $t - 1$ and year t. $\mathbf{T}_{it}$ is a vector of indicators measuring student i's teacher assignment in year t. β_{it} represents the effect of individual teachers (i.e., teacher value-added) on student learning, and it will be used in the next stage of analysis to estimate teacher performance growth over time. X_{it} includes (a) whether or not a student repeated a grade in year t, (b) his FRPL eligibility, (c) sex, (d) race/ethnicity, (e) whether or not he is classified as gifted, (f) special education status by the type of disability (speech/language disability, learning disability, cognitive/mental disability, physical disability, emotional disability, and other types of disabilities), (g) school mobility, and (h) grade level. We differentiate between two types of school mobility: structural school change and non-structural school change. Structural school change is defined as when at least 30% of student i's classmates from the previous school moved to the same receiving school in the current year. Otherwise, a student school change is defined as non-structural.

Student test scores are normalized by year and grade so that they have a mean of 0 and a standard deviation of 1. One concern with the gains model is that score gains are often higher for students who start at a lower initial performance level. This correlation could be the result of regression to the mean; it could also result from the properties of state-designed standardized tests, which may have more differentiation power at the lower end of the student ability distribution than at the higher end. Consequently, the value-added scores of teachers in high-performing classrooms and schools, estimated using state standardized tests, could be penalized. Following a strategy suggested by Hanushek, Kain, O'Brien, and Rivkin (2005), we divide students into deciles according to their lagged test scores and then standardize score gains within each lagged score decile.

Estimating Teacher Performance Trajectories

Next, we use the estimated teacher value-added as the outcome in estimating individual teachers' performance trajectories as they gain experience. Value-added scores are all centered on within-year averages. In other words, a teacher's performance is always measured relative to all other teachers in a particular year. Therefore, the estimated performance trajectory reflects how a teacher's *relative* performance changes with experience. Our assumption is that the average teacher performance remains roughly constant from year to year.

As pointed out in the introduction, early work on productivity returns to experience uses cross-sectional data that do not truly track the same teachers over time. Rather than capturing within-teacher productivity growth, those experience–productivity profiles reflect differences between teachers with varying experience levels instead.

With the performance of the same teacher measured across multiple time points, we can estimate teacher-specific performance trajectories using a three-level HLM. We formulate our Level 1 model as a measurement model, which emphasizes that value-added scores are *estimates* of "true" teacher performance. At Level 2, teacher-by-year performance is modeled as a function of within-teacher variation in experience and other time-varying classroom and school characteristics. The intercept and the performance–experience slope of the Level 2 equation are allowed to vary randomly across teachers at Level 3. Specifically, we estimate the following model:

Level 1 (Measurement Model):

$$\mathrm{b}_{tj} = \beta_{tj} + \varepsilon_{tj}. \tag{2}$$

Level 2 (Teacher-by-Year):

$$\begin{aligned}\beta_{tj} = {} & \pi_{0j} + \pi_{1j}\left(Experience_{tj}\right) \\ & + \pi_{2j}\left(Experience_{tj}^{2}\right) + Z_{tj}^{'}\pi_{qj} + e_{tj}.\end{aligned} \tag{3}$$

Level 3 (Teachers):

$$\begin{aligned}\pi_{0j} &= \beta_{00} + r_{0j} \\ \pi_{1j} &= \beta_{10} + r_{1j} \\ \pi_{2j} &= \beta_{20} + r_{2j} \\ \pi_{qj} &= \beta_{q0}.\end{aligned} \tag{4}$$

Substituting Equations 2 to 4 into one another successively yields the equivalent single-equation mixed-effects model:

$$\begin{aligned}b_{tj} &= \beta_{00} + \beta_{10}\left(Experience_{tj}\right) + \beta_{20}\left(Experience_{tj}^2\right) \\ &\quad + Z_{tj}'\beta_{q0} + r_{0j} + r_{1j}\left(Experience_{tj}\right) \\ &\quad + r_{2j}\left(Experience_{tj}^2\right) + e_{tj} + \varepsilon_{tj}.\end{aligned} \tag{5}$$

Level 1 is a measurement model where a teacher j's value-added score in year t, b_{tj}, is an estimate of his or her true performance β_{tj}. Because a teacher's performance is estimated based on the test scores of a student sample drawn from the larger student population, one of the components of variation in the estimated teacher value-added arises from sampling error. The sampling error associated with each value-added score is assumed to be normally distributed as $\varepsilon_{tj} \approx N(0, V_j)$. As standard error is estimated for each teacher value-added score from our first-stage analysis, ε_{tj} can be represented by these standard error estimates and the sampling variance can be assumed known. Raudenbush and Bryk (2002) label this type of model as "*level-1 variance-known* (or V-known)" models, which are widely used in meta-analysis studies but are also generally applicable to research problems where "a single statistic (say a standard deviation, proportion, or correlation) is available from each of many contexts and the goal is to compare these statistics" (Raudenbush & Bryk, 2002, p. 207).

At Level 2, teacher j's true performance in year t (β_{tj}) is modeled as a function of experience and other time-varying covariates $\boldsymbol{Z}_{tj}$. For novice teachers, for instance, the intercept π_{0j} estimates the value-added for teacher j when he or she had 1 year of experience. It randomly varies across teachers in the Level 3 model, with a grand mean of β_{00} and a variance of VAR(r_{0j}) ≡ τ_{00}. The slope π_{1j} estimates teacher-specific performance trajectories or returns to experience. It represents the instantaneous rate of growth in the first time period of each corresponding teacher career stage. Like the intercept, π_{1j} is also allowed to randomly vary across teachers at Level 3, with a mean return to experience of β_{10} and a variance of VAR(r_{1j}) ≡ τ_{11}. The coefficient π_{2j}, similarly a random variable, captures possible non-linearity in teacher performance trajectories.

We estimate this three-level model for three subgroups of teachers separately: novice teachers, early-career teachers, and mid-career teachers. For each group of teachers, we track their value-added scores for up to 5 years. For early-career teachers and mid-career teachers, they start out in our samples with 6 and 11 years of experience, respectively. We recenter the experience variable accordingly, such that π_{0j} always corresponds to teacher value-added in those starting years. Similarly, β_{1j} represents the instantaneous rate of growth in experience year 6 and 11.

We could alternatively have kept all teachers in one sample and estimate teacher performance growth trajectories at once. However, very few teachers (less than 2% of all teachers) have data points spanning all the years of our study period and therefore individual teachers' growth rates at one career stage or another would have to be extrapolated. Although our technical ability to fit a growth model is not affected by missing data, such extrapolation is credible only when data are missing at random (see Singer & Willett, 2003, for a discussion on various types of "missingness"). In other words, when we fit a multilevel model for change, we have to implicitly assume that each teacher's observed records are a random sample of data from his or her underlying true growth trajectory. This assumption is likely to be untenable, as teacher attrition and mobility are demonstrated to be associated with teacher performance (e.g., Feng & Sass, 2011; Goldhaber, Gross, & Player, 2007; Hanushek et al., 2005).

By dividing teachers into subgroups of similar career stages, we increase the chance of teachers in each subgroup having complete data and avoid the need to extrapolate mid-career growth trajectories for teachers who are only observed during their earlier years (and vice versa). Teacher performance growth is allowed to vary by career stages, and within each career stage, our model allows for quadratic growth trajectories. This strategy is likely to permit sufficient flexibility to capture non-linearities that are substantively important.

Finally, because our value-added model did not control for any school effects or classroom characteristics, changes in teacher effectiveness over time may reflect changes in the overall school quality. Teachers may also be assigned to different types of students and classrooms as they gain seniority. If seniority-related teacher–classroom sorting patterns differed systematically between high- and lower-poverty schools, estimated differences in teacher performance trajectories between school types would be confounded. To mitigate these concerns, our HLM model includes additional Level 2 control variables $\boldsymbol{Z}_{tj}$. These include the average value-added of a teacher's peers in the same school,[2] the classroom average pretest scores, and the standard deviation of pretest scores within classrooms.

Compare Teacher Performance Trajectories

In the final stage of our analysis, we use the estimated performance trajectory for each individual teacher as the dependent variable and explore how teacher growth varies both between high- and lower-poverty school settings and within each type of school-poverty setting. Variation in teacher performance trajectories could be associated with a number of factors in addition to school poverty. Kraft and Papay (in press) and Loeb, Beteille, and Kalogrides (2012), for instance, identify perceived school working conditions and overall school effectiveness as factors related to teacher growth. With the data available to us, we are limited to exploring three possible correlates of teacher growth and their interaction with school poverty: teachers' initial value-added, the number of school changes a teacher made during the study period, and the effectiveness gap between a teacher and her peers in the starting years of each career stage. We make three hypotheses:

Hypothesis 1: Teachers who started highly effective may have flatter performance trajectories (either because they have less room to grow, or because of "regression-to-the-mean").

Hypothesis 2: Frequent switches from one school to another may slow a teacher's growth as he or she frequently needs to relearn his or her working environment.

Hypothesis 3: Having colleagues more effective than oneself may help one develop faster.

More importantly to our current context, these factors may differ systematically along the school-poverty dimension, helping partially explain any observed teacher growth differentials by school-poverty settings. Therefore, we compare teacher performance trajectories between high- and lower-poverty schools after controlling for these factors (and interacting them with school-poverty level) in a regression analysis. As the dependent variable (teacher performance trajectory) is an estimate rather than a known value, we employ a feasible generalized least squares (FGLS) method developed by Lewis and Linzer (2005) to account for estimation error.

Findings

At the start of each career stage, teachers who teach consistently in high-poverty schools for the next 5 years appear to be statistically different from those who persist in lower-poverty schools (Table 2). For example, in both states, elementary school mathematics teachers always staffed in high-poverty schools are more likely to start with temporary or provisional licenses than teachers in lower-poverty schools. The difference is more apparent among novice teachers than teachers with 5 to 6 or 10 to 11 years of experience. In addition, in comparison with novice teachers in high-poverty schools, novice teachers in lower-poverty schools are more likely to start their teaching careers with graduate degrees. In North Carolina (where data are available), novice teachers in lower-poverty schools are more likely to be certified through traditional state accredited education programs. In contrast, nearly half of novice teachers in high-poverty schools are certified through non-traditional routes to teaching, such as lateral entries, licensing through regional alternative licensing centers, and teaching permits obtained under the state's Alternative Entry regulations.

Beyond these differences in background characteristics, novice teachers' value-added scores vary depending on where they began their teaching careers. For example, novice teachers in

TABLE 2

Base Year Characteristics of Elementary School Mathematics Teachers, by School Setting and Teacher Career Stages

	North Carolina		Florida	
	Always in lower-poverty schools	Always in high-poverty schools	Always in lower-poverty schools	Always in high-poverty schools
Novice teachers				
Regular license (%)	96.14	89.11**	97.46	95.37**
Graduate degree (%)	11.71	8.36**	21.25	17.99**
Traditional route (%)	57.67	51.00**		
Praxis score (*SD*)	0.33	0.12**		
Value-added scores (*SD*)	−0.03	−0.05*	−0.09	−0.11**
Observations	2,408	909	3,765	3,704
Early-career teachers				
Regular license (%)	99.31	98.03**	99.96	99.84
Graduate degree (%)	22.27	17.30**	31.00	32.34
Traditional route (%)	60.37	63.15		
Praxis score (*SD*)	0.20	−0.10**		
Value-added scores (*SD*)	0.01	0.00	0.044	0.044
Observations	1,877	659	2,545	1,945
Mid-career teachers				
Regular license (%)	99.58	98.82*	99.87	99.89
Graduate degree (%)	28.42	28.20	38.46	38.02
Traditional route (%)	67.18	72.45**		
Praxis score (*SD*)	0.11	−0.26**		
Value-added scores (*SD*)	0.02	0.02	0.061	0.059
Observations	1,418	422	1,560	881

Note. Schools with 60% or more students eligible for free/reduced-price lunch are defined as high-poverty schools. Otherwise, they are defined as lower-poverty schools.
*Statistically significant at .10. **Statistically significant at .05.

high-poverty schools tend to start with significantly lower value-added than those in lower-poverty schools. In both states, the teacher effectiveness gap is about 0.02 standard deviations of student gain scores. By comparison, at the start of the early- and mid-career stages, teachers in high- and lower-poverty school settings are equally effective on average (Figure 1 depicts the distribution of teacher value-added in the initial year of each corresponding career stage). However, PRAXIS test scores, an indicator of teacher aptitude and available in the North Carolina data, are significantly lower among teachers in high-poverty schools than those in lower-poverty schools across all career stages. The average gaps are sizable and range from 0.20 to 0.37 standard deviations.

These baseline differences between teachers in high- and lower-poverty schools are important to keep in mind when we examine variation in teacher performance trajectories: Not only may the rate of improvement be related to where a teacher starts, but also growth is more imperative for teachers who start at a lower level than for teachers who are already highly effective.

Table 3 reports the estimated random and fixed effects corresponding to novice, early-career, and mid-career teacher samples. The top panel reports variance components in an unconditional HLM model with no covariates. For teachers at all career stages in North Carolina, slightly more than 50% of the total (net of variance due to estimation error) variation in teacher value-added is between teachers. The remaining

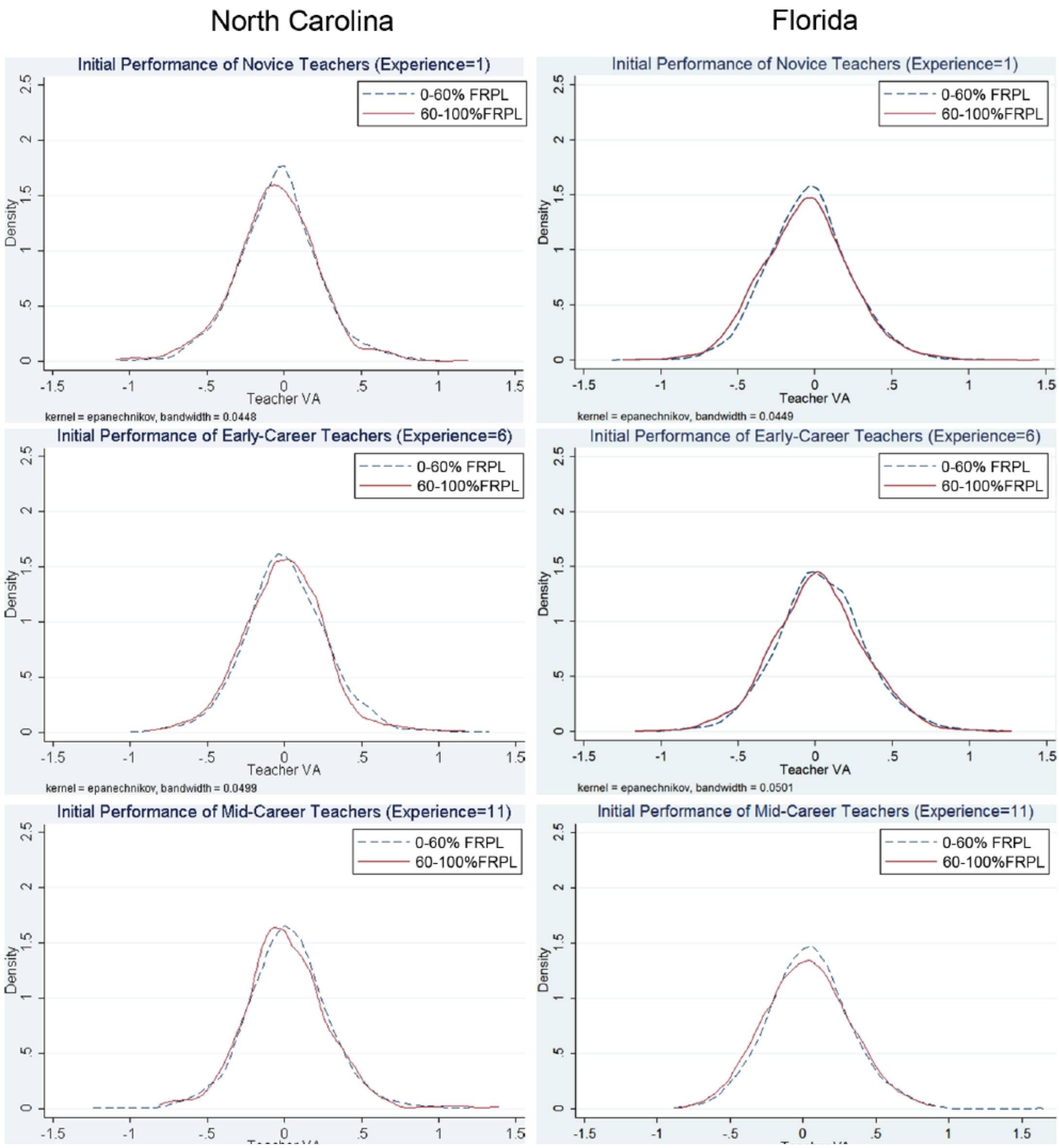

FIGURE 1. *The distribution of elementary school math teachers' VA, by state, year of experience, and school-poverty setting.*
Note. FRPL = free/reduced-price lunch; VA = value-added.

variation is attributed to within-teacher changes over time. About one third of the year-to-year within-teacher value-added fluctuations, however, can be explained by teacher experience as well as annual changes in classroom characteristics and teacher–peer performance. For all teacher groups, 1 standard deviation difference in teacher value-added is estimated to be associated with 0.24 (square root of the between-teacher variance component) standard deviations in student mathematics score gains.

The size of the variance components among elementary school mathematics teachers in Florida is very comparable with that in North Carolina. About 50% to 60% of the total (net of variance due to estimation error) variation in

TABLE 3

Variance Components and Fixed-Effects Estimates for Elementary School Mathematics Teachers, by Experience Level, State, and Model

	Novice teachers		Early-career teachers		Mid-career teachers	
	North Carolina	Florida	North Carolina	Florida	North Carolina	Florida
Unconditional model						
Variance component						
Between teacher, τ_{00}	.060**	.071**	.058**	.080**	.061**	.075**
Within teacher, σ^2	.059**	.070	.053**	.055	.055**	.055
ICC (τ_{00} / (τ_{00} + σ^2))	.50	.50	.52	.59	.53	.58
Model with Level 2 covariates						
Variance component						
Between teacher, τ_{00}	.056**	.059**	.077**	.078**	.073**	.077**
Random effects variance for return to experience, τ_{11}	.011*	.011**	.012**	.004	.004	.002
Random effects variance for return to experience^2, τ_{22}	.000	.000*	.000**	.000	.000	.000
Within teacher, σ^2	.039**	.049**	.035**	.041**	.040**	.042**
Fixed effect						
Mean initial value-added, β_{00}	−.122**	−.147**	−.008	.055**	.009	.042*
	(.010)	(.014)	(.012)	(.019)	(.014)	(.025)
Mean return to experience, β_{11}	.074**	.066**	−.003	−.009	.015*	.011
	(.007)	(.007)	(.007)	(.008)	(.008)	(.010)
Mean return to experience^2, β_{22}	−.009**	−.007**	.001	.002	−.002	−.003
	(.001)	(.001)	(.001)	(.002)	(.002)	(.002)
Classroom average pretest score, β_{33}	−.119**	−.023**	−.103**	−.018*	−.142**	−.041**
	(.011)	(.008)	(.012)	(.010)	(.016)	(.013)
Classroom *SD* of pretest scores, β_{44}	.028**	.049**	.039**	−.023	.029**	.011
	(.010)	(.015)	(.011)	(.020)	(.012)	(.027)
Average peer teacher value-added, β_{66}	.659**	.812**	.562**	.732**	.546**	.734**
	(.023)	(.022)	(.028)	(.028)	(.032)	(.030)

Note. Fixed effects presented with robust standard errors. The measurement model at Level 1 takes into account estimated standard errors associated with teacher value-added estimates.
*Statistically significant at .10. **Statistically significant at .05.

yearly teacher value-added is between teachers. The remaining variation, of which around one fourth can be explained by experience gains and other time-varying classroom and school factors, is year-to-year changes within teachers. One standard deviation difference in teachers' value-added is associated with 0.27 standard deviations in student mathematics score gains at all career stages.

The mean annual performance growth rate among novice teachers is 0.07 standard deviations in student mathematics score gains both states (Table 3). However, teacher performance growth starts to slow down even during the first 5 years of teaching. This is captured by the negative and statistically significant coefficients on the quadratic form of experience. Among early-career teachers, the mean performance trajectory is flat and statistically insignificant. However, among mid-career teachers with at least 10 or 11 years of experience, teacher performance starts to improve again in North Carolina (although at a much slower rate of 0.015 standard deviations in student score gains), but remains flat in Florida. On average, teacher

TABLE 4

Distribution of Performance Trajectories Related to Experience Among Elementary School Mathematics Teachers, by State, Experience Level, and School-Poverty Setting

	North Carolina			Florida		
	25th percentile	50th percentile	75th percentile	25th percentile	50th percentile	75th percentile
Novice teachers	.051	.074	.096	.044	.065	.086
Lower-poverty schools	.052	.075	.097	.044	.065	.086
High-poverty schools	.049*	.073	.093**	.045	.066	.086
Early-career teachers	−.024	−.004	.017	−.020	−.009	.003
Lower-poverty schools	−.024	−.004	.018	−.021	−.009	.003
High-poverty schools	−.024	−.004	.016	−.020	−.009	.003
Mid-career teachers	.004	.015	.026	.003	.011	.019
Lower-poverty schools	.004	.015	.026	.003	.011	.019
High-poverty schools	.005	.014	.025	.003	.010**	.018

Note. Schools with 60% or more students eligible for free/reduced-price lunch are defined as high-poverty schools. Otherwise, they are defined as lower-poverty schools.
*Statistically significant at .10. **Significantly different from lower-poverty school estimates at the .05 level.

performance does not appear to decline as teachers become more experienced. In other words, there is no evidence of teacher "burnout."

The distinct teacher performance trajectories at various career stages are consistent with findings reported in the literature on teacher experience and performance. However, average growth rates mask significant variation in teacher performance improvement across individuals, especially at the novice level. In both states, the variation in teacher growth rate is statistically significant. Among novice teachers in North Carolina, slower-growth teachers (those whose growth rate is at the 25th percentile) improve their performance by 0.05 standard deviations in student score gains annually, which is still a considerable rate of growth. However, the faster-growth teachers (those with growth rate at the 75th percentile) improve their performance by more than 0.09 standard deviations, almost 80% faster than their slower-improving peers (Table 4). In other words, faster-growing teachers gain more than half a year equivalent of average teacher performance growth annually than slower-growth teachers (on average, 1 year of additional experience is associated with 0.07 standard deviations among novice teachers). Similarly, among novice elementary mathematics teachers in Florida, teachers at the top quartile improve their performance by 0.09 standard deviations, compared with 0.04 standard deviations among teachers at the bottom quartile.

Compared with novice teachers, the variation in performance trajectory among early-career teachers is smaller in both states. It remains significant in North Carolina and but becomes statistically insignificant in Florida (Table 3). At this career stage, it appears that the median teacher in both states has stopped improving, while the performance of teachers near the bottom quarter starts to decline (at an annual rate of −0.02 standard deviations or worse). However, the performance of some teachers continues to improve in North Carolina, even though at a slower rate than the growth rate during the first 5 years of teaching.

Throughout the performance trajectory distribution among mid-career teachers, few teachers appear to experience a substantial amount of performance decline. Even at the bottom quartile of the performance trajectory, teacher performance trajectories largely remain flat in both states. At the top quartile, teachers grow at a moderate rate of 0.03 standard deviations in student score gains annually in North Carolina and 0.02 standard deviations in Florida. This is consistent with recent empirical findings that teachers with 10 or more years of experience can continue to improve.

Considerable variation in teacher performance trajectories, however, does not appear to be related to school student poverty rates. Table 4 shows no significant difference in teacher growth between teachers in high- and lower-poverty schools throughout the distribution of performance trajectories. However, as demonstrated earlier, teachers in high-poverty schools start at a lower performance level than teachers in lower-poverty schools. As growth rates are likely to be inversely correlated with starting performance levels, we compare teacher growth trajectories in high- and lower-poverty schools using multiple regressions that control for teacher's starting performance level and interact it with school-poverty status.

In addition, some teachers in our samples may have switched between schools with similar student poverty rates during the 5-year study period. Such school switches may affect a teacher's performance trajectory because of potential disruptions to development every time he or she changes schools. If teachers in one type of schools tended to be more mobile than other teachers, comparisons of average teacher performance trajectories between school settings could be biased. As a result, our regressions also control for the number of school switches a teacher made.

Finally, as reported in Tables 3, a significant amount of year-to-year fluctuation in a teacher's value-added is explained by the average quality of her colleagues, consistent with the findings documented by Jackson and Bruegmann (2009). However, it is unclear whether the association between a teacher's own value-added and her colleagues' is because having higher quality colleagues makes an average teacher improve faster, or because teachers tend to work with other teachers with similar performance levels. In other words, it is not clear whether the observed association is the result of growth or sorting. Feng and Sass (2011) find that a teacher is more likely to leave a school or exit teaching when the gap between her performance and that of her peers is large, possibly supporting the theory of teacher self-sorting into groups of similar performance levels. To directly explore whether the relative quality of a teacher's peers is associated with how fast he or she improves, our regression models include a variable that measures the average difference between a teacher's value-added and the value-added of her colleagues, calculated at the start of each career stage, and its interaction with school-poverty status.

The results of these regressions are reported in Tables 5 to 7. Overall we still find that teacher performance trajectories do not differ significantly by school-poverty settings. Novice teachers in North Carolina's high-poverty schools and mid-career teachers in Florida's high-poverty schools appear to grow slightly slower than their counterparts in lower-poverty schools. In both cases, the statistically significant differences are substantively small.

Contrary to our hypothesis, teacher performance trajectories are not always negatively related with starting performance levels. In both states, it is the higher-performing novice teachers who improve the fastest (Table 5). In the first 5 years of teaching, it appears that the initial gap in teaching effectiveness could further widen. This pattern holds in both high- and lower-poverty school settings, and it seems to be consistent with the findings reported by Atteberry, Loeb, and Wyckoff (2013) that the initially lowest-performing teachers on average fail to "catch up" with other teachers and remain consistently the lowest performing even after 5 years.

The relationship between teachers' initial performance levels and performance growth rates starts to diverge between North Carolina and Florida among early- and mid-career teachers. Whereas better teachers continue to improve at faster rates than teachers with lower starting performance level in Florida, the opposite is true among North Carolina teachers (Tables 6 and 7). Similar contradiction is found in the direction of the coefficients on the initial performance gap between oneself and peers. In Florida, positive coefficients indicate that teachers improve faster when they are more effective than their peers in the same school; by contrast, North Carolina teachers appear to improve faster when they have colleagues better than themselves. Reconciling these contradictions between the two states is beyond the scope of this study. However, they warrant further investigation.

Next, we divide teachers by their starting performance level and explore whether high-performing (top quarter based on starting year value-added)

TABLE 5

Relationship Between Teacher Growth Trajectories and Covariates Among Novice Teachers, by State and Sample

	North Carolina			Florida		
	All teachers	Low-performing teachers	High-performing teachers	All teachers	Low-performing teachers	High-performing teachers
High FRPL school	−0.003*	0.003	−0.009	−0.001	−0.002	0.004
	(0.002)	(0.008)	(0.007)	(0.001)	(0.005)	(0.004)
Base year VA	0.047**	0.050**	0.069**	0.035**	0.031**	0.047**
	(0.005)	(0.015)	(0.014)	(0.004)	(0.011)	(0.012)
High FRPL × Base year VA	0.003	0.007	0.016	−0.014**	−0.021	−0.015
	(0.009)	(0.024)	(0.026)	(0.006)	(0.014)	(0.015)
Number of school changes	−0.000	0.006*	−0.006	−0.004**	0.002	−0.010**
	(0.002)	(0.004)	(0.004)	(0.001)	(0.003)	(0.003)
VA gap from peer teachers	0.008	−0.013	0.010	0.021**	0.007	0.040**
	(0.005)	(0.012)	(0.010)	(0.005)	(0.009)	(0.009)
High FRPL × VA gap	0.008	0.028	−0.005	−0.011*	0.020*	−0.007**
	(0.010)	(0.018)	(0.020)	(0.006)	(0.011)	(0.013)
Constant	0.077**	0.072**	0.071**	0.069**	0.064**	0.063**
	(0.001)	(0.005)	(0.004)	(0.001)	(0.003)	(0.003)
Observations	3,308	814	825	6,941	1,729	1,776
R^2	.136	.055	.093	.161	.285	.108

Note. Standard errors in parentheses. All regressions take into account estimation errors in the dependent variable. FRPL = free/reduced-price lunch; VA = value-added.
*p < .10. **p < .05.

and low-performing (bottom quarter) teachers demonstrate distinct growth trajectories in schools with different student poverty rates. Similarly, we find no systematic difference in teacher performance growth between school-poverty settings, regardless of teachers starting performance level. Our findings are best depicted in Figures 2 and 3, where the *y*-axis represents the predicted teacher value-added (based on parameters estimated for the HLM model) and the *x*-axis represents years of experience. The solid lines represent teachers in lower-poverty schools and dashed lines represent teachers in high-poverty schools. The bolded, black lines depict the average predicted value-added by years of experience among *all* teachers in each school-poverty setting. The unbolded, colored lines represent the predicted value-added for subgroups of teachers categorized by their starting performance level (in quarters).

Unlike typical teacher productivity–experience profiles, our figures reflect estimates based on three distinct teacher samples, each representing teachers at a different career stage. We present teacher growth trajectories as disconnected line segments to highlight the fact that we are not tracking the same teachers over 15 years. Figures 2 and 3 visually demonstrate that the growth trajectories for teachers in both high- and lower-poverty schools almost always parallel each other. In addition to this key finding, these two figures also reflect a few other noteworthy patterns. First, we observe some evidence that teacher value-added regresses to the mean (i.e., the initial performance gaps between the most and the least productive teachers narrow in later years), particularly among early- and mid-career teachers. This is not surprising as teacher performance is measured with error. As a result, the initial performance gap, as measured by single-year value-added scores, is likely to overstate the permanent performance difference among teachers. Second, the same measurement error problem

TABLE 6

Relationship Between Teacher Growth Trajectories and Covariates Among Early-Career Teachers, by State and Sample

	North Carolina			Florida		
	All teachers	Low-performing teachers	High-performing teachers	All teachers	Low-performing teachers	High-performing teachers
High FRPL school	−0.001	0.007	−0.012	−0.001	0.010**	−0.001
	(0.002)	(0.008)	(0.008)	(0.001)	(0.004)	(0.004)
Base year VA	−0.018**	−0.056**	−0.024*	0.017**	0.004	−0.020**
	(0.006)	(0.016)	(0.014)	(0.003)	(0.009)	(0.009)
High FRPL × Base year VA	−0.001	0.041	0.038	0.009*	0.033**	0.014
	(0.010)	(0.028)	(0.027)	(0.005)	(0.013)	(0.013)
Number of school changes	−0.002	0.002	−0.003	0.002	0.001	−0.002
	(0.002)	(0.003)	(0.003)	(0.001)	(0.003)	(0.002)
VA gap from peer teachers	−0.028**	−0.005	−0.022**	0.023**	0.023**	0.031**
	(0.006)	(0.011)	(0.011)	(0.004)	(0.007)	(0.007)
High FRPL × VA gap	−0.003	−0.022	−0.026	−0.012**	−0.013	−0.014
	(0.010)	(0.019)	(0.019)	(0.005)	(0.010)	(0.010)
Constant	−0.002**	−0.010**	−0.001	−0.010**	−0.014**	−0.014**
	(0.001)	(0.004)	(0.004)	(0.001)	(0.010)	(0.003)
Observations	2,521	625	630	4,393	1,092	1,115
R^2	.110	.050	.056	.168	.055	.097

Note. Standard errors in parentheses. All regressions take into account estimation errors in the dependent variable. FRPL = free/reduced-price lunch; VA = value-added.
*$p < .10$. **$p < .05$.

(and therefore regression-to-the-mean) should apply to novice teachers as well. Yet despite this factor, the best novice teachers seem to grow faster than the worst, at least in the first 3 years of teaching. And finally, among teachers in all three career stages, those starting in the bottom quarter of the performance distribution fail to catch up with other teachers and remain the lowest performing, in spite of regression-to-the-mean.

It becomes clear that school-poverty settings are not correlated with variations in teacher performance trajectories in any systematic way. However, we define high- and lower-poverty schools at the 60% FRPL cutoff, and some schools on both sides of the cutoff may be very similar in terms of school-poverty settings. To check the sensitivity of our findings, we redefine lower-poverty schools as those enrolling less than 40% FRPL-eligible students, and keep high-poverty schools defined as those with 60% or more FRPL-eligible students. Our findings remain very similar (Table 8). That is, there is generally no systematic difference in teacher performance trajectories between school-poverty settings. Categorizing schools into even more distinct poverty categories (≥70% FRPL vs. <30% FRPL), we still could not find systematic differences of meaningful size in teacher growth.

Summary and Discussion

This study is a descriptive analysis on whether teachers in high-poverty schools and lower-poverty schools follow different performance growth paths. It is directly motivated by Sass et al.'s (2012) study, which finds differential "returns" to experience between teachers in high- and lower-poverty schools. Specifically, they find no evidence of increased productivity among teachers with greater levels of experience in high-poverty schools, whereas the productivity profile among teachers in lower-poverty schools

TABLE 7

Relationship Between Teacher Growth Trajectories and Covariates Among Mid-Career Teachers, by State and Sample

	North Carolina			Florida		
	All teachers	Low-performing teachers	High-performing teachers	All teachers	Low-performing teachers	High-performing teachers
High FRPL school	0.001	−0.003	0.005	−0.001*	−0.009**	−0.001
	(0.001)	(0.006)	(0.006)	(0.001)	(0.003)	(0.004)
Base year VA	−0.011**	0.000	−0.008	0.017**	0.015	0.029**
	(0.004)	(0.012)	(0.010)	(0.003)	(0.006)	(0.007)
High FRPL × Base year VA	0.003	−0.017	−0.007	−0.004	−0.028**	−0.007
	(0.008)	(0.024)	(0.020)	(0.004)	(0.011)	(0.012)
Number of school changes	0.001	0.003	−0.000	0.0003	−0.002**	−0.004
	(0.001)	(0.002)	(0.003)	(0.0001)	(0.003)	(0.003)
VA gap from peer teachers	−0.012**	−0.020**	−0.009	0.010**	0.017**	−0.004
	(0.004)	(0.009)	(0.008)	(0.003)	(0.005)	(0.006)
High FRPL × VA gap	−0.002	0.005	0.001	0.005	0.006	0.013
	(0.007)	(0.017)	(0.014)	(0.004)	(0.008)	(0.009)
Constant	0.015**	0.016**	0.013**	0.010**	0.012**	0.009**
	(0.001)	(0.003)	(0.003)	(0.0003)	(0.002)	(0.002)
Observations	1,837	458	454	2,350	602	587
R^2	.071	.033	.022	.220	.078	.073

Note. Standard errors in parentheses. All regressions take into account estimation errors in the dependent variable. FRPL = free/reduced price lunch; VA = value-added.
$^*p < .10$. $^{**}p < .05$.

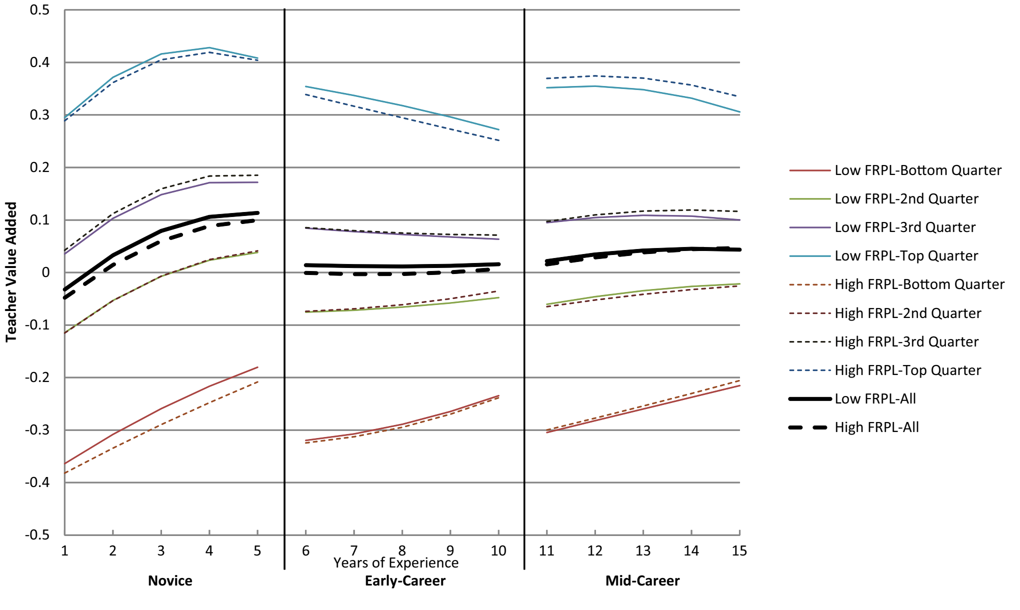

FIGURE 2. *Performance trajectory of elementary school math teachers in North Carolina: By career stage, initial value-added, and school-poverty status.*
Note. FRPL = free/reduced-price lunch.

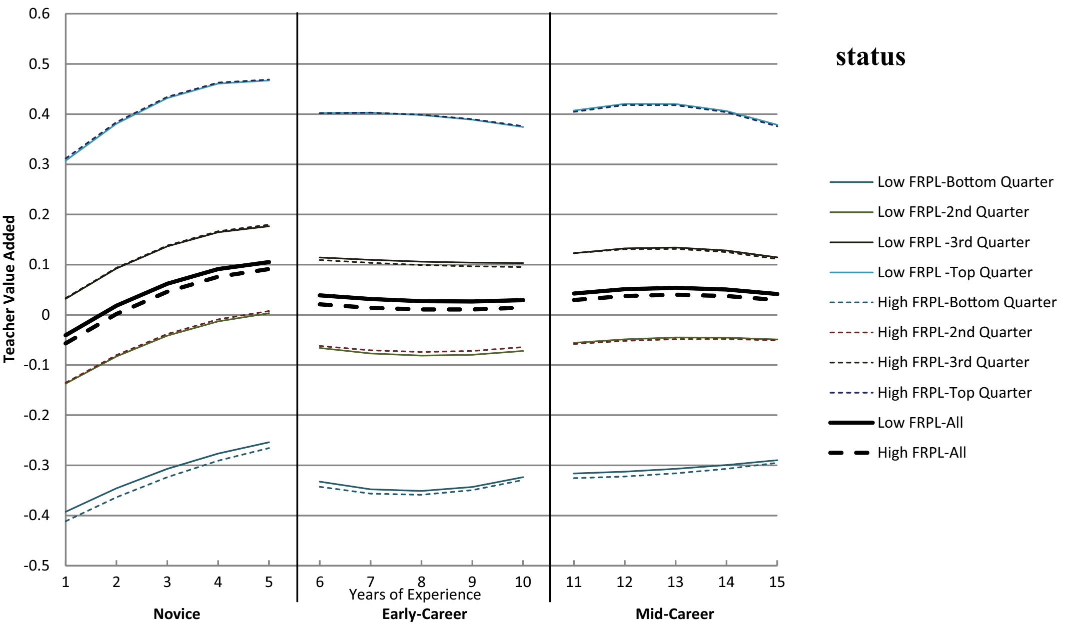

FIGURE 3. *Performance trajectory of elementary school math teachers in Florida: By career stage, initial value-added, and school-poverty status.*
Note. FRPL = free/reduced-price lunch.

TABLE 8
Varying High- and Low-Poverty School Definitions, by State and Teacher Experience Level

	North Carolina		Florida	
Estimated differences in growth rate	≥60% vs. <40% FRPL	≥70% vs. <30% FRPL	≥60% vs. <40% FRPL	≥70% vs. <30% FRPL
Novice teachers	−0.006**	−0.007**	−0.001	−0.002*
	(0.002)	(0.002)	(0.001)	(0.001)
Observations	2,013	1,118	5,225	3,618
Early-career teachers	−0.001	0.003	0.001	0.0003
	(0.002)	(0.002)	(0.001)	(0.001)
Observations	1,563	869	3,328	2,271
Mid-career teachers	0.002	−0.001	−0.002**	−0.001
	(0.001)	(0.002)	(0.001)	(0.001)
Observations	1,116	626	1,767	1,134

Note. Standard errors in parentheses. Regression specifications remain the same as in Tables 5 through 7. FRPL = free/reduced-price lunch.
$^*p < .10$. $^{**}p < .05$.

suggests large growth over the first 5 years and then the rate of growth flattens out in later years.

Because that study is based on cross-sectional data, it is unclear what may be underlying the observed experience–productivity profiles. The authors suggested at least two possible explanations. The first is that teachers learn at different rates in high- and lower-poverty school settings. For instance, teachers in more disadvantaged schools may need to provide more support to their students or devote more time to classroom discipline, which diverts their energy away from perfecting their instruction. The second possible explanation is teacher-school sorting. If less-productive experienced teachers were more likely to move into high-poverty schools, cross-sectional data would also show a flat experience–productivity profile for high-poverty school teachers.

Determining which of these two hypotheses might drive the differential returns to experience has important policy implications. If the teacher learning hypothesis is correct, policymakers may instead promote teacher professional development in high-poverty schools. If the sorting hypothesis is true, our teacher labor policy should probably focus more on the equitable distribution of high- and low-performing teachers.

Our results show that nearly half of the total variation in teachers' yearly value-added estimates (net of variance due to estimation error) is within teachers. About one fourth to one third of these within-teacher year-to-year changes can be explained by teachers' increasing experience level, as well as other time-varying classroom and school-level characteristics. We find teacher-specific performance trajectories to be the steepest among novice teachers. It becomes flat among early-career teachers, but performance improvement resumes among mid-career teachers, at least in the North Carolina data.

The average experience–productivity profiles at all career stages mask significant variations across teachers. Compared with the slowest improving teachers, the fastest improving teachers gain more than half a year equivalent of performance growth annually. Such variations, however, are not systematically correlated with school-poverty status. Therefore, our findings are inconsistent with the teacher learning hypothesis proposed to explain the lack of "return" to experience among teachers in high-poverty schools. In addition, these findings contradict the hypothesis that teacher burnout in high-poverty schools explains differential returns to experience. Rather, the evidence suggests that teacher mobility and attrition patterns across schools may be a more plausible explanation for the inequitable distribution of particularly low-performing teachers in high-poverty schools.

The lack of association between teacher performance growth and school-poverty settings, combined with the finding that novice teachers with lower initial effectiveness also improve more slowly, suggests the importance of recruitment and selective retention. Teacher human capital management policies recommended by some researchers, such as lowering barriers to entry, is likely to exacerbate unequal distribution of effective teachers between high- and lower-poverty schools. However, our findings seem to support other recommendations such as selective retention of teachers based on just their first year of performance.

Finally, we want to emphasize that our findings do not suggest that teacher performance growth cannot be influenced by school-level factors other than school-poverty settings. For instance, Kraft and Papay (in press) report that teachers improve faster in schools that are perceived as more supportive. Loeb et al. (2012) report that teachers who work in schools that were more effective at raising achievement in a prior period improve more rapidly in a subsequent period than do those in less-effective schools. What our findings suggest is that even though school poverty is often a key focus of education policies and interventions, it is not a deciding factor of teacher performance growth. Future research and teacher labor market policies should shift their focus away from school-poverty status to other dimensions of school environments that may be conducive to teacher growth.

Authors' Note

The views expressed are those of the authors and should not be attributed to the American Institutes for Research, its trustees, or any of the funders or supporting organizations mentioned herein. Any errors are attributable to the authors.

Acknowledgments

The authors gratefully acknowledge comments and feedback from Mike Garet, Jane Hannaway, Kilchan Choi, Mengli Song, Kennan Cepa, and anonymous reviewers for this journal.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Bill & Melinda Gates Foundation. They thank the North Carolina Education Research Data Center and the Florida Department of Education's K–20 Education Data Warehouse (through the National Center for the Analysis of Longitudinal Data in Education Research, supported by Grant R305A060018 to the American Institutes for Research from the Institute of Education Sciences, U.S. Department of Education) for providing access to data.

Notes

1. We have also used the level-score as the dependent variable and controlled for the lagged score and its quadratic term on the right-hand side of the equation. Resulting value-added estimates are very highly correlated with estimates from the gain score model at .96.

2. Past research has shown that a teacher's colleagues play a significant role in his or her productivity (Jackson, 2012; Jackson & Bruegmann, 2009). Moreover, peer quality may be indicative of a school's ability in attracting, developing, and retaining good teachers.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, *25*, 95–135.

Atteberry, A., Loeb, S., & Wyckoff, J. (2013, February). *Do first impressions matter? Improvement in early career teacher effectiveness* (Working Paper No. 90). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research (CALDER).

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008, May). *Who leaves? Teacher attrition and student achievement* (Working Paper No. 14022). Cambridge, MA: National Bureau of Economic Research.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*, 2633–2679.

Clotfelter, C. T., Ladd, H., & Vigdor, J. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed-effects. *Journal of Human Resources*, *45*, 656–681.

Feng, L., & Sass, T. (2011, January). *Teacher quality and teacher mobility* (Working Paper No. 57). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research (CALDER).

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *Review of Economics and Statistics*, *89*, 134–150.

Goldhaber, D., Gross, B., & Player, D. (2007, October). *Are public schools really losing their best? Assessing the career transitions of teachers and their implications for the quality of the teacher workforce* (Working Paper No. 12). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research (CALDER).

Goldhaber, D., & Hansen, M. (2010). Using performance on the job to inform teacher tenure decisions. *American Economic Review*, *100*, 250–255.

Gordon, R., Kane, T., & Staiger, D. (2006, April). *Identifying effective teachers using performance on the job* (Hamilton Project white paper). Washington, DC: Brookings Institution.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005, February). *The market for teacher quality* (Working Paper No. 11154). Cambridge, MA: National Bureau of Economic Research.

Hanushek, E. A., Kane, T., & Rivkin, S. (2004). Why public schools lose teachers. *Journal of Human Resources*, *39*, 326–354.

Hanushek, E. A., & Rivkin, S. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, *100*, 267–271.

Jackson, C. K. (2010, May). *Match quality, worker productivity, and worker mobility: Direct evidence from teachers* (Working Paper No. 15990). Cambridge, MA: National Bureau of Economic Research.

Jackson, C. K. (2012, October). *Do high-school teachers really matter?* (Working Paper No. 17722). Cambridge, MA: National Bureau of Economic Research.

Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, *1*(4), 85–108.

Koedel, C., & Betts, J. (2007, April). *Re-examining the role of teacher quality in the educational production function* (Working paper). Nashville, TN: National Center on Performance Incentives, Peabody College, Vanderbilt University.

Kraft, M. A., & Papay, J. P. (in press). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*.

Lewis, J. B., & Linzer, D. A. (2005). Estimating regression models in which the dependent variable is based on estimates. *Political Analysis*, *13*, 345–364.

Loeb, S., Beteille, T., & Kalogrides, D. (2012). Effective schools: Teacher hiring, assignment, development, and retention. *Education Finance and Policy*, *7*, 269–304.

Miller, R., & Chait, R. (2008). *Teacher turnover, tenure policies, and the distribution of teacher quality: Can high-poverty schools catch a break?* Washington, DC: Center for American Progress.

National Council on Teacher Quality. (2012). *State of the States 2012: Teacher effectiveness policies* (NCTQ State Teacher Policy Yearbook Brief). Washington, DC: National Council on Teacher Quality.

Papay, J. P., & Kraft, M. A. (2011). *Do teachers continue to grow with experience?* [Mimeo]. Harvard University, Cambridge, MA.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: SAGE.

Rockoff, J. (2004). The impact of individual teachers on students' achievement: Evidence from panel data. *American Economic Review*, *94*, 247–252.

Sass, T., Hannaway, J., Xu, Z., Figlio, D., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, *72*, 104–122.

Scafidi, B., Sjoquist, D., & Stinebrickner, T. (2007). Race, poverty, and teacher mobility. *Economics of Education Review*, *26*, 145–159.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford, UK: Oxford University Press.

Staiger, D., & Rockoff, J. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives*, *24*(3), 97–118.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, *113*, F3–F33.

Wiswall, M. (2013). *The dynamics of teacher quality. Journal of Public Economics*, *100*, 61–78.

Xu, Z., Hannaway, J., & Taylor, C. (2011). Making a difference? The effect of Teach for America on student performance in high school. *Journal of Policy Analysis and Management*, *30*, 447–469.

Xu, Z., Ozek, U., & Corritore, M. (2012, June). *Portability of teacher effectiveness across school settings* (Working Paper No. 77). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research (CALDER).

Authors

ZEYU XU is a principal researcher in the CALDER Center at the American Institutes for Research. His research interests include teacher labor market issues, randomized-controlled trial designs, and standards-based educational reforms.

UMUT ÖZEK is a principal researcher in the CALDER Center at the American Institutes for Research. His research interests include school choice, school accountability, and immigration.

MICHAEL HANSEN is a principal researcher in the CALDER Center at the American Institutes for Research. His research focuses primarily on teacher productivity and the teacher labor market.

Manuscript received March 27, 2014
Revision received September 12, 2014
Accepted October 22, 2014