Is the Persistence of Teacher Effects in Early Grades Larger for Lower–Performing Students?
Author(s): Spyros Konstantopoulos and Min Sun
Reviewed work(s):

# Is the Persistence of Teacher Effects in Early Grades Larger for Lower-Performing Students?

SPYROS KONSTANTOPOULOS
*Michigan State University*

MIN SUN
*Virginia Tech*

We examined the persistence of teacher effects from grade to grade on lower-performing students using data from Project STAR. Teacher effects were computed as residual classroom achievement within schools. Teacher effects in one grade predicted achievement in following grades using quantile regression. Results consistently indicated that all students benefited similarly from teachers, and differential teacher effects were not evident. Overall, lower-performing students benefit as much as other students from teachers except in fourth grade, where lower-performing students benefit more. Having effective teachers in successive grades seems beneficial to lower-performing students in mathematics and reading. However, having low-effective teachers in successive grades is detrimental to all students especially in mathematics.

Since the Coleman report, much educational research has focused on identifying school-related factors that affect student learning, and many school policy initiatives have attempted to ensure that school resources are allocated adequately to schools. One factor widely believed by educational researchers to affect student achievement is teachers, and a fundamental goal of teacher effects research is to examine how teachers improve academic achievement for all students. Since the US educational system is designed to provide all students with equal access to school resources and to reduce inequality in achievement, it is important to determine whether lower-performing students benefit more from teachers than other students. It is appealing to think that teachers increase academic achievement for all students and simultaneously close the achievement gap between higher- and lower-performing students by

helping lower-performing students perform as well as higher-performing students.

One focus of No Child Left Behind (NCLB) has been to reduce the achievement gap and to ensure that lower-performing students from disadvantaged backgrounds attain academic proficiency. One important mechanism through which this can be accomplished is teacher effectiveness. NCLB has mandated state plans to improve teacher effectiveness, with the underlying belief that effective teachers can increase achievement especially for lower-performing students. It is timely then to examine how teachers affect lower-performing students and whether these types of students benefit more from teachers.

Anecdotal evidence as well as empirical research indicates that teachers differ noticeably in their effectiveness as educators and pedagogues to promote student achievement. Evidence from experimental and nonexperimental studies has consistently indicated that teachers differ considerably in their effectiveness and that teacher effects are large (e.g., Goldhaber and Brewer 1997; Nye et al. 2004; Rivkin et al. 2005; Rowan et al. 2002). In these studies, teacher effectiveness is defined typically as differences or variation in achievement between classrooms adjusted by student background.

Findings about differential teacher effects on minority and disadvantaged students have been mixed. For example, some researchers have demonstrated that teacher characteristics such as experience are positively and significantly linked to the achievement of black students (Murnane and Phillips 1981). Other researchers have shown that teacher effects are not associated with the achievement of black or Hispanic students (Hanushek 1992). More recent work has reported that minority and disadvantaged students seem to benefit as much as other students from teachers (Konstantopoulos 2009). Other recent work has provided evidence about the persistence of teacher effects in elementary grades (Konstantopoulos and Chung 2011). Specifically, the authors reported that teacher effects were positive and persisted through sixth grade. However, the differential persistence of teacher effects on lower-performing students has not been well documented.

In this study, we examined the persistence of the effects teachers have on

SPYROS KONSTANTOPOULOS is associate professor and program coordinator of measurement and quantitative methods in the Department of Counseling, Educational Psychology, and Special Education at Michigan State University. His research interests include teacher and school effects, class size effects, power analysis in experimental designs with nested data structures, and multilevel meta-analysis. MIN SUN is an assistant professor of Educational Leadership and Policy Studies at Virginia Tech. Her research focuses on educational policy analysis, the development and evaluation of teachers and principals, and quantitative research methodology.

lower-performing students from grade to grade using high-quality experimental data from Project STAR (Student Teacher Achievement Ratio; Krueger 1999; Nye et al. 2000). Specifically, we were interested in investigating the differential persistence of teacher effects in early grades across the achievement distribution in order to determine whether lower-performing students in one grade benefit more from teacher effects in the previous grade. Project STAR was a well-executed, large-scale randomized experiment, and evidence derived from such data is likely to have higher internal validity and, to a lesser extent, higher external validity than small-scale studies with convenience samples. We used quantile regression to compute the persistence of teacher effects across the entire distribution of achievement. The outcome variables were mathematics and reading scores, and the main independent variable was teacher effects.

## Differential Teacher Effects

The computation of the persistence of teacher effects at different quantiles of the achievement distribution allowed us to detect possible differential effects. Such effects indicate that certain groups of students are affected by teachers differently than other students and that the effectiveness of teachers varies by achievement level. When differential effects are evident, the changes in achievement for lower- and higher-performing students that are due to teachers vary. A notion related to differential effects is that of interaction effects between teacher effects and levels of achievement. The idea is that teacher effects may interact with levels of achievement, and through that interaction the effects are potentially maximized. The notion of interaction effects between variables goes back to the pioneering work of Cronbach and Snow (1977). Such effects indicate the degree to which teacher effects depend on the level of achievement.

In our study, prior teacher effects were used to predict future performance of lower-performing students. One hypothesis is that lower-performing students may benefit more from having effective teachers than other students. Alternatively, the performance of such students may be influenced more by teachers and less by parents. For example, effective teachers may be more likely to identify lower-performing students and provide instruction that is designed to benefit these students in the early grades. If that were true, then it is also likely that the persistence of these effects will be larger for lower-performing students the following year. Alternatively, instructional practices enacted by effective teachers may engage or motivate lower-performing students more in learning activities, and such gains may persist from year to year.

*Persistence of Teacher Effects*

We used quantile regression to estimate teacher effects at different quantiles of the achievement distribution. These estimates indicate the degree of interaction between the persistence of teacher effects and level of achievement. When the estimates are significant, the research hypothesis that teacher effects vary by level of achievement is tenable, and the null hypothesis that effects are similar for all students is false. In particular, we were interested in whether lower-performing students in one grade (e.g., first grade) benefited more from having effective teachers in the previous grade (e.g., kindergarten). If that hypothesis were true, one would expect larger estimates of the persistence of teacher effects in the lower tail of the achievement distribution.

In the context of answering the question *what works for whom*, such analyses can yield important findings. First, the analyses can reveal the groups of students that benefit more from having effective teachers in previous grades. The magnitude of the estimates suggests whether the persistence of teacher effects is considerable enough to be of policy relevance. Second, knowing whether all or some students benefit similarly or differently from having had effective teachers is potentially valuable. Such information provides an explicit inference about the generality of the results and points to the consistency of the persistence of teacher effects for different groups of students.

## Previous Research on Teacher Effects

Generally, there are two major lines of research that have discussed the effects of teachers on student achievement. The first tradition of research includes studies that measure the association between teacher characteristics and student achievement. The second tradition of research estimates the variation in achievement between classrooms.

### Teacher Characteristics and Student Achievement

Three areas are included in this tradition of research. The first area includes education production function studies that attempt to determine the relationship between specific, measured teacher characteristics such as teacher experience, education, salary, or certification and student achievement. However, because parents choose neighborhoods in which to live and hence their associated schools, according to tastes and resources, student background is confounded with teacher characteristics (Tiebout 1956). Therefore, education production function studies attempt to control for this confounding by using student background characteristics as covariates in regression models (e.g., Coleman et al. 1966). A particularly important covariate is prior achievement,

because it summarizes the effects of individual background. Some reviewers of the education production function literature argue that measured teacher characteristics such as educational preparation, experience, or salary are only slightly related to student achievement (Hanushek 1986). Other reviewers argue that some of the resource characteristics such as teacher experience and teacher education have positive effects on student achievement (Greenwald et al. 1996).

More recently, researchers have examined the effects of teacher experience, knowledge, and certification on student achievement. Economists have demonstrated a positive association between teacher experience and student achievement (Clotfelter et al. 2006). Education researchers have examined the effect of teacher content knowledge on student achievement (Hill et al. 2005; Kennedy 2008). For instance, Hill and colleagues found that teachers' mathematical knowledge was a significant and positive predictor of mathematics achievement gains in first and third grades, controlling for student socioeconomic status (SES) and teacher characteristics such as experience. Kennedy (2008) also found that teacher content knowledge seems to benefit students. Finally, researchers have provided evidence that National Board–certified teachers seem to be more effective than other teachers (Goldhaber and Anthony 2007).

The second area includes studies known as process-product studies that aim to identify classroom processes (e.g., observed teacher characteristics and teaching practices) that are associated with student outcomes (or products) such as achievement (Good and Brophy 1987). In these studies, teacher confidence in teaching students successfully, efficient allocation of classroom time to instruction and academic tasks, effective classroom organization and group management, and active/engaging teaching that emphasizes understanding of concepts have been shown to affect student achievement positively (Good and Brophy 1987). Reviewers of teacher effects from process-product studies have concluded that effective teachers substantially influence academic achievement for all students (Good 1979). In addition, other studies have documented that teachers with higher evaluation scores in their teaching had higher classroom achievement means and contributed in closing the achievement gap between lower- and higher-SES students in some grades (Borman and Kimball 2005). Improvements in teacher qualifications also seem to increase student achievement, especially in poor schools (Boyd et al. 2008).

The third area is known as value-added research. Value-added models have gained considerable attention in the last 15 years, mainly because of the urgency of using achievement scores to determine teacher effects on student outcomes following the passage of NCLB. The underlying idea in value-added models is to examine the effects of teachers on students' learning gains net of student background. These models intend to estimate the unique contribution

or "value added" of teachers on students' change in learning. In this sense, value-added research is not a completely new area since regression models that examine teacher and school effects net of student background date back to the famous Coleman report.

Meyer (1997) argues that the key objective in value-added research is to determine teacher effects on student achievement net of the effects of other sources that may affect student achievement. It is common practice in value-added research to gauge teacher effects via regression models that control for covariates hypothesized to influence student learning such as previous achievement. Often the outcome in such regression models is a posttest measure of student achievement in standardized tests. The main independent variable represents teacher effects, and other variables such as prior measures of student achievement are included as covariates to adjust for previous ability (McCaffrey et al. 2004; Raudenbush 2004). However, not all value-added models control for student background. For example, some researchers who have used value-added models have argued that sometimes controlling for student background may overadjust the teacher effects estimates (Ballou et al. 2004; Sanders and Rivers 1996).

In principle, value-added models are hypothesized to provide more accurate and perhaps causal estimates of teacher effectiveness than other studies. However, value-added models do not necessarily eliminate all possible confounding effects because unobservables may still be related to teacher effects (Braun 2005). Rubin et al. (2004) provided a thoughtful discussion about causal inferences of teacher and school effects. In their discussion, the authors argued that causal estimates of teacher effects are difficult to conceptualize even in well-done randomized experiments and that value-added models do not necessarily provide causal estimates. Rather, these models should more likely be conceptualized as providing descriptive measures of teacher effects. Along this line, more recent work has raised concerns about the assumptions that underlie value-added models and has proposed that these models should go through rigorous validation and falsification tests (Rothstein 2010).

Previous work has also examined the persistence of teacher effects on student achievement using value-added models (e.g., Ballou et al. 2004; McCaffrey et al. 2004; Sanders and Rivers 1996). For example, Sanders and Rivers (1996) used a value-added model to predict the teacher effects in grades 3, 4, and 5 on fifth-grade achievement, controlling for achievement in second grade. The authors concluded that the teacher effects were cumulative. More recent work has also demonstrated that teacher effects persist in elementary grades and that their cumulative effects are considerable (Konstantopoulos and Chung 2011).

*Studies of Variation in Teacher Effects*

The second tradition of research examines the variation in achievement between classrooms controlling for student background. These models typically use prior achievement as a covariate as well and measure the variance in residualized student achievement gain across classrooms. That is, these classroom variances in achievement gain are due to differences in teacher effectiveness. The underlying assumption is that the between-classroom variation in achievement is caused by variation in teacher effectiveness. Typically these studies calculate the proportion of variance in residualized student achievement gain accounted for by teacher effects using regression analysis. Specifically, the change in the coefficient of determination is estimated when teacher effects are included in the regression model, and this change indicates the variability in achievement across classrooms due to teachers. Overall, the results of such studies have suggested that there is indeed considerable variation in teacher effectiveness (Goldhaber and Brewer 1997; Murnane and Phillips 1981; Rowan et al. 2002). Further, a recent study documented large differences in average achievement among classrooms (Nye et al. 2004). Nye et al. reviewed teacher effects estimates in the literature and suggested that, on average, a one standard deviation increase in teacher effectiveness would increase student achievement gains by about one-third of a standard deviation. A more recent review summarized estimates of teacher effects in standard deviation units and reported that in reading the estimates ranged from approximately one-tenth to one-fifth of a standard deviation and in mathematics from one-tenth to one-third of a standard deviation (Hanushek and Rivkin 2010).

One caveat, however, about the studies within this tradition of research is that they cannot identify specific teacher characteristics that compose teacher effectiveness. It is noteworthy that typically observed teacher characteristics such as teacher experience and education explain only a small proportion of the variation in teacher effectiveness (Konstantopoulos 2011b; Rivkin et al. 2005). For example, Konstantopoulos found that teacher education and experience explained less than 1 percent of the variation in teacher effects in early grades. These findings suggest that the majority of the variability in teacher effects remains unexplained and not captured by typically observed teacher characteristics. It is possible that the teacher characteristics typically measured are easy to collect but unrelated to achievement, whereas other characteristics such as teacher motivation remain unmeasured because they are difficult to collect. Even if researchers attempted to measure the "right" teacher characteristics, it is possible that the measurement would be so poor that the effects would be attenuated.

## Limitations of Previous Work

Students are frequently assigned to teachers on the basis of their characteristics such as achievement; teachers are not randomly assigned to classrooms either. For instance, more experienced teachers may be assigned to classes composed of higher-performing students as a privilege of seniority or to classes composed of lower-performing students as a compensatory strategy. This nonrandom assignment creates problems when inferring the relation between teacher characteristics and student achievement because the causal direction of the relationship is unclear. In a recent study, Clotfelter et al. (2006) reported that economically advantaged students are more likely to have highly qualified teachers than other students. This finding suggests that, owing to selection, the association between teacher characteristics and achievement can be biased.

Because of confounding, it is difficult to interpret the estimates of teacher effects on student achievement in both traditions of research mentioned above. Although it is essential to control for student background in order to reduce variability in preexisting differences and identify the unique contribution of teachers to student achievement, even important covariates such as prior achievement and SES do not completely eliminate differences in background characteristics. Teacher effects may still be confounded with unobserved individual, family, school, and neighborhood variables. For example, previous achievement or SES may not adequately control for preexisting differences in unobservables such as motivation.

The problems in interpretation would be eliminated if both students and teachers were randomly assigned to classes. Random assignment of students would in principle ensure that all observable and unobservable differences between students in different classes are not systematic. Likewise, random assignment of teachers to classes would assure that any differences in teacher characteristics are uncorrelated with classroom achievement and other classroom variables (Weiss 2010). In this study, we used data from Project STAR that satisfy both conditions of random assignment. Project STAR was a field experiment designed to measure class size effects. However, the fact that students and teachers were randomly assigned to classroom types within schools in each grade provides a great opportunity to gauge teacher effects since the potential confounding issues should in principle be reduced if not eliminated.

Validity of Project STAR

*Random Assignment*

The internal validity of the Project STAR estimates depends on whether random assignment effectively eliminated preexisting differences between students and teachers assigned to different types of classrooms. The fact that the random assignment of students and teachers to classrooms was carried out by a consortium of researchers enhances its credibility. However, it is good practice to check for preexisting differences of observed characteristics of teachers and students. Unfortunately, no pretest scores were collected in Project STAR, so it was not possible to examine differences in prekindergarten achievement. However, one could check the degree to which random assignment was successful using student variables such as age, race, and SES. Krueger (1999) examined the success of random assignment among treatment groups (i.e., small, regular, and regular classes with a full-time aide) and found that in observed variables such as SES, minority group status, and age, there were no significant differences between classroom types once school differences were taken into account. Krueger also found that there were no significant differences across classroom types with respect to teacher characteristics such as race, experience, and education. He concluded that random assignment did not seem to be compromised. Other analyses however, have raised some concerns about the reliability of random assignment, especially for variables such as age and SES (Hanushek 1999; Konstantopoulos 2011a).

Even if we assume that random assignment across classroom types was successful, it is still possible that classrooms assigned to the same treatment group within schools were different. Because teacher effects are computed using differences in average achievement between classrooms that receive the same treatment type within schools, it is critical to check whether random assignment was successful across classrooms within treatment types within schools. A recent study undertook that task and produced results that are consistent with what would be expected had random assignment been successful. That is, no systematic differences were found for observed student characteristics between classrooms that were in the same treatment type within schools (see Nye et al. 2004).

*Attrition*

Most large-scale longitudinal studies such as Project STAR suffer from attrition. Approximately 28 percent of the students who participated in Project

STAR in kindergarten were not part of the study in the first grade. The attrition rate was nearly 25 percent for students who participated in the study in the first grade but were not present in the second grade. Twenty percent of the students dropped out of the study after the second grade, and thus they did not participate in the third grade. Across all grades, about 50 percent of the students who were part of the experiment in kindergarten were still part of Project STAR in the third grade. Thirty-eight percent of the students who were part of the experiment in kindergarten were still in the study in the fourth grade.

The effects of differential attrition on the estimates of class size have been discussed in two studies (Krueger 1999; Nye et al. 2000). It is common practice to examine differential attrition between types of classrooms on the outcome measures such as achievement scores. For example, Krueger examined whether differential attrition among types of classrooms biased the estimates of class size. Differential attrition can bias class size effects if the students who dropped out of small classes were systematically different in achievement from those who dropped out of regular-type classes (Krueger 1999). In longitudinal designs such as Project STAR, one way to measure the effects of differential attrition is to impute the scores of those students who dropped out of the study each year (Krueger 1999). Krueger computed the class size estimates with and without imputation, compared the estimates, and concluded that it is unlikely that differential attrition biased the class size estimates. The same conclusion was reached by Nye et al. (2000) independently using slightly different methods. Nonetheless, recent analyses have suggested some evidence that attrition was related to school achievement and school composition, that is, the proportion of minority or disadvantaged students (Konstantopoulos 2011a). In this article we attempted to adjust for possible selection from one grade to the next by using the Heckman method, as illustrated in the analysis section (Heckman 1979).

## Methodology

### Data

Project STAR is a four-year large-scale experiment that was conducted in Tennessee in the mid-1980s. The experiment was commissioned in 1985 by the Tennessee state legislature and was implemented by a consortium of universities and the Department of Education in Tennessee. The experiment lasted for four years, from kindergarten to third grade, and the total cost, including hiring teachers and teacher aides, was about $12 million. The State

of Tennessee paid to hire additional teachers and classroom aides. Project STAR is considered one of the greatest experiments in education.

In the first year of the experiment, a cohort of more than 6,000 kindergartners in more than 300 classrooms of 79 elementary schools in 42 Tennessee districts participated. The sample included a broad range of schools and districts, for example, urban, rural, wealthy, and poor. Districts had to agree to participate for four years; allow school visits for verification of class sizes, interviews, and data collection; and include extra student testing. They also had to allow research staff to assign pupils and teachers randomly to class types and to maintain the assignment of students to class types from kindergarten through third grade.

Kindergarten students were assigned randomly to different types of classrooms within each school: small classes (with 13–17 students), regular classes (with 22–26 students), or regular classes with a full-time aide. Teachers were also assigned randomly to classes of different types. The students who entered the study in the first, second, or third grade were assigned randomly to classes at that time. Teachers at each grade were also assigned randomly to classes as the experimental cohort passed through the grades.

## Statistical Analysis

*Computing teacher effects in each grade.*—The main objective of our study was to examine whether teacher effects in one year (e.g., kindergarten) are associated with different levels of achievement (e.g., low, medium, high) in the following year (e.g., first grade). First, we computed teacher effects within each grade. This analysis makes use of the Stanford Achievement Test (SAT-9) reading and mathematics test scores collected as part of Project STAR. SAT-9 is a widely used test that measures academic achievement of elementary and secondary school students. Owing to the random assignment of students and teachers to classrooms within schools, the classrooms within each school should be initially equivalent, and hence, any systematic differences in achievement among classes must be due to one of two sources: the class size effect or differences in teacher effectiveness. Thus, within each school, any systematic differences in achievement between classrooms that had the same treatment must be due to differences in teacher effectiveness (see Nye et al. 2004).

Following Nye et al. (2004), we operationalize teacher effects as classroom-specific residuals or random effects. The variance of these random effects indicates the magnitude of the effects (see Nye et al. 2004). Because the data were produced from a randomized experiment in which students and teachers were randomly assigned to classrooms within schools, it is likely that the confounding between student background and teacher effects is reduced or

minimized and that these classroom residuals may represent the "true" teacher effects (Raudenbush 2004).

Teacher effects were adjusted for treatment effects (e.g., class size) and possible student effects (e.g., age, gender, race, and SES) or classroom context effects (e.g., peer effects). It is crucial to adjust for class size effects because it is likely that class size plays a role in achievement differences between classrooms; for example, smaller classes may have higher achievement than other classes. Similarly, classroom context could contribute to achievement differences between classrooms. For example, differences in the proportion of minority or low-SES students between classrooms could explain part of the achievement differences between classrooms. One way to model classroom context in statistical models is to use variables that represent peer effects. We included in our model peer effects for variables such as gender, race, SES, and age (see the specification in eq. [1] below). A typical way of modeling peer effects is to compute aggregate measures of student variables for each classroom within each school for all students in the classroom except a specific student (see Ammermueller and Pischke 2009; Mashburn et al. 2009). To illustrate the process, suppose that there are 21 students in a classroom. To compute the peer effect index for the twenty-first student with respect to low SES, we computed the number of the remaining 20 students in the classroom who were eligible for free or reduced-price lunches and then divided that number by 20. If 10 of the 20 students were eligible for free or reduced-price lunches in the classroom, then the peer effect for the twenty-first student is $10/20 = 0.50$. That is, the peer effects are classroom averages, but they are computed for each student separately.

Finally, student characteristics such as minority and SES could play a role in achievement differences between classrooms. For example, the average achievement of a classroom may be higher because of the proportion of high-SES students in the classroom and not necessarily because of the effectiveness of the teacher. As a result, typically, student variables are included as covariates in statistical models that measure teacher or school effects. It is difficult to know whether that proportion of the between-teacher variance explained by student variables is attributed solely to student background. Similarly, it is difficult to know whether the proportion of the between-teacher variance explained by student variables should be considered a teacher effect. Because of this uncertainty, we decided to follow a conservative approach and estimated teacher effects controlling for student background. We acknowledge that there is a possibility that student background is confounded with teacher effects. We also acknowledge that our assumption that the between-teacher variance explained by student variables is due solely to student background may not hold exactly. If the between-teacher variance explained by student variables includes teacher effects, then our teacher effects estimates are underestimated

because the distribution of teacher effects has a smaller variance. Empirically, the proportion of between-teacher variance explained by student background variables ranged between 3 and 5 percent in kindergarten and first grade and between 7 (in mathematics) and 14 (in reading) percent in second and third grades. In mathematics, the proportion of variance explained is very small overall and should not affect our estimates that much. In reading, however, the proportion of variance explained is a little larger, especially in third grade.

We computed teacher effects as classroom-specific random effects or residuals by employing a three-level model (Bryk and Raudenbush 1988). The first level involves a between-student within-classroom-and-school model, the second level involves a between-classroom within-school model, and the third level is a between-school model. To compute teacher effects, we used the same specification for mathematics and reading achievement for each grade separately. Hence, for each grade, the one-level regression equation for student $i$ in class $j$ in school $k$ is

$$
\begin{aligned}
Y_{ijk} = {} & \beta_{000} + \beta_{100}\text{FEMALE}_{ijk} + \beta_{200}\text{LOWSES}_{ijk} + \beta_{300}\text{MINORITY}_{ijk} \\
& + \beta_{400}\text{AGE}_{ijk} + \beta_{500}\text{PEERFEM}_{ijk} + \beta_{600}\text{PEERLOWSES}_{ijk} \\
& + \beta_{700}\text{PEERMINORITY}_{ijk} + \beta_{800}\text{PEERAGE}_{ijk} + \beta_{010}\text{SMALL}_{jk} \\
& + \beta_{020}\text{AIDE}_{jk} + \varepsilon_{ijk} + \xi_{0jk} + \eta_{00k},
\end{aligned}
\tag{1}
$$

where $Y_{ijk}$ represents student achievement in mathematics or reading; FEMALE is a dummy variable for gender; LOWSES is a dummy variable for free or reduced-price lunch eligibility; MINORITY is a dummy variable for minority group membership (more than 90 percent are African Americans); AGE represents students' age; the PEER variables indicate peer effects for female, low SES, minority, and age, respectively; SMALL is a dummy variable for being in a small class; AIDE is a dummy variable for being in a regular class with a full-time teacher aide; $\varepsilon_{ijk}$ is a student-specific random effect; $\xi_{0jk}$ is a classroom-specific random effect; and $\eta_{00k}$ is a school-specific random effect. All random effects are normally distributed with zero means and constant variances. The $\beta$'s are the regression coefficients across all students, classrooms, and schools. For simplicity, all predictors were fixed, and only the classroom-specific and school-specific intercepts were treated as random at the second and third levels, respectively. In this model, the variance of the error term is divided into three parts: the within-classroom, the between-classroom within-school, and the between-school variance. The classroom-specific random effects, $\xi$'s, represent the teacher effects adjusted for student, peer, and class size effects. In this analysis, we used intention to treat (ITT) assignment to classes to control for class size effects. The ITT is unbiased by design and does not incorporate any possible validity threats that may have occurred during the experiment (Freedman 2006).

We also computed teacher effects in the first, second, and third grades using a slightly different specification that included prior achievement (see eq. [2] below). The model illustrated in equation (2) incorporated prior achievement and prior classroom achievement. We call these teacher effects residualized teacher effects. The overwhelming majority of studies that measure peer effects do not include current-grade classroom achievement in their models because of the reflection problem identified by Manski (1993). Because of the reciprocal nature of the determination of peer achievement, this peer component, that is, the current-grade aggregate classroom achievement, is likely endogenous and differentiated from the aggregate classroom measures of family background (Hanushek et al. 2003). Specifically, student and peer achievements are determined simultaneously, and therefore current-grade classroom achievement is likely endogenous. By and large, when researchers estimate peer effects, they use reduced forms (see eq. [1]) that do not include current-grade aggregate classroom achievement (Ammermueller and Pischke 2009). Some researchers, however, have used lagged peer achievement (e.g., previous achievement) when they estimate peer effects, but the endogeneity problem may still hold (Hanushek et al. 2003; Lefgren 2004). It is unclear which model is best when estimating peer effects. In our study, the objective was to sort out peer effects from teacher effects. To that end and because of the longitudinal nature of the data, we were able to include lagged peer achievement when we estimated teacher effects in the first, second, and third grades. For example, in the first grade, we computed teacher effects as second-level residuals controlling for several variables as well as achievement in kindergarten and lagged peer achievement in kindergarten. For each grade (i.e., first, second, or third) the model becomes

$$
\begin{aligned}
Y_{ijk} = {} & \beta_{000} + \beta_{100}\text{FEMALE}_{ijk} + \beta_{200}\text{LOWSES}_{ijk} + \beta_{300}\text{MINORITY}_{ijk} \\
& + \beta_{400}\text{AGE}_{ijk} + \beta_{500}\text{PRACHIEVEMENT} + \beta_{600}\text{PEERFEM}_{ijk} \\
& + \beta_{700}\text{PEERLOWSES}_{ijk} + \beta_{800}\text{PEERMINORITY}_{ijk} \\
& + \beta_{900}\text{PEERAGE}_{ijk} + \beta_{100}\text{PEERPRACHIEVEMENT} \\
& + \beta_{010}\text{SMALL}_{jk} + \beta_{020}\text{AIDE}_{jk} + \hat{\varepsilon}_{ijk} + \hat{\xi}_{0jk} + \hat{\eta}_{00k},
\end{aligned}
\tag{2}
$$

where PRACHIEVEMENT indicates previous achievement (e.g., mathematics) and PEERPRACHIEVEMENT indicates lagged peer achievement. The $\beta$'s are the regression coefficients across all students, classrooms, and schools.

Because the teacher-specific residuals are computed separately from the school-level residuals, differences in achievement among teachers/classrooms within types of classrooms and within schools should be net of school differences in achievement. That is, the variance of the second-level residuals is the

variance in classroom achievement within treatment types and within schools adjusted for school effects expressed as variability in achievement between schools in the third-level residuals.

*Modeling teacher effects in the following grade.*—The teacher effects (i.e., $\xi$'s) were computed for kindergarten, first, second, and third grades and used as predictors of student achievement in subsequent grades, that is, first, second, third, and fourth. In this analysis, teacher effects were used as a predictor of achievement in the following grade, and the estimates indicate whether the effectiveness of the teacher that a student had in one year persisted and affected that student's achievement in the following year. This analysis used samples of students who were part of Project STAR for two consecutive grades, and it was conducted in two stages as discussed below. To compute the persistence of residualized teacher effects, we used samples in the second, third, and fourth grades.

Since students who stayed in the experiment from grade to grade may be different from those who left the experiment, we tried to adjust for potential selection. One way to control for selection directly in a regression model is to use the Heckman model (Heckman 1979). This model involves two steps. In the first step, we used probit to model whether a student stayed in the experiment from grade to grade. For example, we modeled the probability that a student who participated in the experiment in kindergarten would also participate in the first grade. The binary outcome variable is staying in the study or dropping out. The predictors were carefully chosen to accurately determine the probability that students would stay in the study. The probit model therefore is

$$\text{probit}(\pi) = \beta_0 + \mathbf{XB}, \tag{3}$$

where $\beta_0$ is the constant, and $\mathbf{X}$ is a matrix of variables such as small or regular-size class (ITT), SES and minority status, gender, school urbanicity, and teacher variables such as teacher education, experience, and race. From the probit model we calculated the inverse Mills ratio or lambda ($\lambda$), which we included as a covariate in the second-stage achievement regressions to adjust for possible nonrandom selection of students. The vector $\mathbf{B}$ includes the regression estimates of the probit model. Hence the teacher effects estimates in the achievement regressions were corrected for potential selection. To ensure proper identification of the models, the specification used in the achievement regressions was sufficiently different from those in the probit model; that is, some variables were included in the achievement regression but not in the probit regression.

The second stage of the analysis involved computing teacher effects across the distribution of achievement, for example, lower, middle, and upper tails. To that end, we used quantile regression to estimate teacher effects at various points of the achievement distribution in grades 1–4 (see Buchinsky 1998;

Koenker and Bassett 1978). Education researchers frequently examine the effects of school resources or school interventions on lower-performing, minority, and disadvantaged students. For the purposes of this article, it is possible that teachers in one grade have differential effects on average, lower-, and higher-performing students in the following grade. If all students benefit from teachers equally, then all estimates must be positive and similar in magnitude. If lower-performing students benefit more from teacher effects than other students, then the estimates in the lower tail of the achievement distribution must be larger. Examining the effects of teachers across the entire achievement distribution provides crucial information about reducing the achievement gap. The typical regression model is inadequate to examine the effects of predictors at different points (called quantiles) of the outcome distribution, and as a result we used quantile regression (Hao and Naiman 2007).

Quantile regression is a natural extension of the typical linear regression because it estimates how predictors (e.g., teacher effects) affect outcomes (e.g., achievement) not only in the middle but in the tails of the outcome distribution as well. Hence, quantile regression estimates provide a more complete picture of the effects of predictors on the entire distribution of outcomes (Hao and Naiman 2007). Quantile regression is also a more robust method, compared to typical regression, for analyzing skewed distributions with outliers. Currently, quantile regression is a widely used method in economics and social sciences. We argue that this method can also be useful in education research that focuses on educational inequality and the academic prosperity of students especially in the lower tail of the achievement distribution. The purpose of the present study was to determine whether the persistence of teacher effects produces additional benefits in achievement for lower-performing students in grades 1–4. We believe that quantile regression is well suited for this purpose because it shows how teachers in one year can affect the achievement of lower-performing, average, and higher-performing students the following year. In addition, covariate effects are also modeled across the achievement distribution. The same index (e.g., standard deviation units) can be computed for teacher effects on achievement across the entire distribution, and hence, the results across different points (quantiles) of the achievement distribution are comparable.

We ran quantile regressions for mathematics and reading test scores separately for each grade. In each grade, mathematics and reading scores were regressed on teacher effects and other covariates. For example, first-grade mathematics test scores were regressed on teacher effects in kindergarten controlling for covariates in first grade. The regression equation at each quantile is

$$Y_i = \beta_0 + \beta_1 TE_i + \beta_2 \lambda_i + \mathbf{ST}_i \mathbf{B}_3 + \mathbf{CL}_i \mathbf{B}_4 + \mathbf{SC}_i \mathbf{B}_5 + \varepsilon_i, \qquad (4)$$

where $Y$ is mathematics or reading scores; $\beta_0$ is a constant; $TE$ is the teacher

effect in the previous grade; λ represents the sample selection; **ST** is a row vector of student characteristics such as gender, race/ethnicity, and low SES; **CL** is a row vector of classroom/teacher characteristics such as type of classroom (small or regular with aide), teacher race, education, and experience; **SC** is a row vector of school urbanicity indicators and school composition, that is, percentage of minority or disadvantaged students; and ε is the error. The vectors **B** include the regression coefficients that need to be estimated. The estimate of the teacher effect ($\beta_1$) is the most important for this study. For grades 1–3, ITT was used in the equations to control for class size effects. In the fourth grade, the actual class size was utilized. We were not able to use teacher characteristics as covariates in the fourth grade because such data were not available. We examined the teacher effects at the lower tail (tenth and twenty-fifth quantiles), the middle tail (fiftieth quantile), and the upper tail (seventy-fifth and ninetieth quantiles) of the achievement distribution. Since students are nested within classrooms and schools, our data have a nested structure, and it was important to take into account this nesting when computing the standard errors of the regression coefficients. We used STATA to run quantile regression and computed robust standard errors for the quantile regression estimates via the cluster command. The robust standard errors we obtained take into account the clustering nature of the data as well as heteroscedasticity, that is, nonconstant variation. We also ran models to determine whether the teacher effects are nonlinear. Specifically, we ran models that included a quadratic term for teacher effectiveness as well as the linear term.

*Teacher effects across grades.*—Previous work has examined the cumulative nature of teacher effects on student achievement using value-added models (e.g., Ballou et al. 2004; McCaffrey et al. 2004; Sanders and Rivers 1996). For example, Sanders and Rivers (1996) used a value-added model to predict the teacher effects in grades 3–5 on fifth-grade achievement, controlling for achievement in second grade. The authors found that teacher effects were cumulative. In the same vein, we ran additional analysis to examine whether the effects of teachers through grades are cumulative. One hypothesis is that these effects could be more pronounced for lower-performing students who have had more effective teachers in successive grades. This analysis included all students who were in the study for five consecutive years, kindergarten through fourth grade. We defined more effective teachers as those who were in the top half of the teacher effects distribution in each grade (e.g., kindergarten through third grade). Less effective teachers were defined as those who were in the bottom half of the teacher effects distribution in each grade. We coded the cumulative effects of less or more effective teachers as binary indicators that took the value of one if a student had low- or high-effective teachers in all four years, that is, kindergarten through third grade, and zero otherwise. We then used quantile regression and regressed fourth-grade math-

ematics or reading achievement on the cumulative teacher effects and other covariates, for example, gender, race, SES, and class size in fourth grade. We computed estimates of teacher effects in the tenth, twenty-fifth, fiftieth, seventy-fifth, and ninetieth quantiles. Following this, we coded cumulative teacher effects using the top half or bottom half coding scheme because only a small proportion of students received very ineffective or very effective teachers in consecutive grades. The quantile regressions were

$$Y_i = \beta_0 + \beta_1 \text{TOP50}_i + \beta_2 \text{CLSIZE}_i + \mathbf{ST}_i\mathbf{B}_3 + \mathbf{SC}_i\mathbf{B}_5 + \varepsilon_i \qquad (5)$$

or

$$Y_i = \tilde{\beta}_0 + \tilde{\beta}_1 \text{BOTTOM50}_i + \tilde{\beta}_2 \text{CLSIZE}_i + \mathbf{ST}_i\tilde{\mathbf{B}}_3 + \mathbf{SC}_i\tilde{\mathbf{B}}_5 + \tilde{\varepsilon}_i,$$

where TOP50 (or BOTTOM50) indicates teacher effects from kindergarten through third grade, and CLSIZE represents classroom size in fourth grade. All other variables have been defined previously. The vectors **B** include the regression estimates that need to be estimated. The most important coefficients in this analysis were $\beta_1$ and $\tilde{\beta}_1$. The remaining betas were the estimates of the covariates.

## Results

### Descriptive Statistics

In kindergarten through third grade, nearly 50 percent of the students were female and had a low SES (see table 1). Approximately one-third of the students were minorities. Twenty-five percent of students were in small classes in grades 1–3. About 80 percent of students had white teachers, and 35 percent of students had teachers with graduate degrees in kindergarten through second grade. The average teacher experience ranged between 9 and 14 years. Approximately 30 percent of students attended inner-city or urban schools, whereas the majority of students attended suburban or rural schools. In the fourth grade, nearly 50 percent of students in the sample were female, about 40 percent were eligible for free or reduced lunches, and 20 percent were minorities. Nearly 85 percent of the students attended suburban or rural schools, whereas only about 15 percent of the students attended inner-city or urban schools. The outcomes of interest were mathematics and reading scores, which were standardized to have a mean of zero and a standard deviation of one. The teacher effects that were computed in each grade were teacher/classroom-specific residuals (see eq. [1]), and as a result they had a mean of zero and standard deviations that indicated the magnitude of the teacher effects (see Nye et al. 2004).

*Descriptive Statistics of Variables in the Samples*

| | GRADE | | | | |
|---|---|---|---|---|---|
| VARIABLE | Kindergarten | First | Second | Third | Fourth |
| Female (%) | 48.62 | 47.96 | 48.30 | 47.99 | 48.53 |
| Minority (%) | 33.03 | 33.41 | 35.22 | 33.71 | 20.11 |
| Low SES (%) | 48.44 | 51.35 | 51.61 | 50.54 | 37.90 |
| Small class (%) | 30.04 | 26.14 | 25.56 | 26.49 | . . . |
| Teacher race: black (%) | 16.50 | 17.48 | 20.37 | 20.87 | . . . |
| Teacher has graduate degree (%) | 34.66 | 34.57 | 37.32 | 44.15 | . . . |
| Teacher experience in years | 9.26 | 11.63 | 13.15 | 13.93 | . . . |
| Inner-city school (%) | 22.58 | 20.21 | 21.65 | 19.63 | 7.49 |
| Urban school (%) | 8.98 | 9.17 | 7.05 | 7.45 | 8.34 |
| Suburban school (%) | 22.32 | 23.22 | 24.99 | 25.29 | 24.51 |
| Rural school (%) | 46.12 | 47.40 | 46.30 | 47.62 | 59.67 |
| Sample size: | | | | | |
| Students | 6,325 | 6,829 | 6,840 | 6,802 | 4,352 |
| Teachers | 325 | 339 | 340 | 336 | 222 |
| Schools | 79 | 76 | 75 | 75 | 62 |

NOTE.—SES is socioeconomic status.

## *Variance Decomposition*

During the first stage of the analysis (see eqq. [1] and [2]), we were able to estimate the variance decomposition in mathematics and reading scores at each level as a percentage of the total variance in the outcomes. For this computation we used two different models: an unconditional model and a full model with covariates. In mathematics in kindergarten through third grade, the student-level variance was between 70 and 74 percent of the total variance, the teacher-level variance was between 11 and 13 percent of the total variance, and the school-level variance was between 15 and 18 percent of the total variance in models with no predictors. When student characteristics, class size, and peer effects were included in the model, the student-level variance was between 65 and 68 percent, the teacher-level variance was between 10 and 11 percent, and the school-level variance was between 8 and 15 percent of the total variance. In fourth-grade mathematics, the student-level variance was 80 percent and the teacher- and school-level variances were each 10 percent of the total variance in models with no predictors. When previous achievement and lagged peer achievement were added to the models in kindergarten through third grade, the variances at the second and the third levels changed slightly.

*Persistence of Teacher Effects*

In reading in kindergarten through third grade, the student-level variance was between 72 and 80 percent of the total variance, the teacher-level variance was between 9 and 11 percent of the total variance, and the school-level variance was between 11 and 18 percent of the total variance in models with no predictors. When student characteristics, class size, and peer effects were included in the model, the student-level variance was between 66 and 71 percent, the teacher-level variance was between 6 and 10 percent, and the school-level variance was between 4 and 14 percent of the total variance. In fourth-grade reading, the student-level variance was 83 percent, the teacher-level variance was 7 percent, and the school-level variance was 10 percent of the total variance in models with no predictors. When previous achievement and lagged peer achievement were added to the models in kindergarten through third grade, the variances at the second and third levels changed slightly.

*Linear Teacher Effects*

The main objective of our study was to examine whether teacher effects persisted from grade to grade and if they were distributed uniformly across the achievement distribution or if there was evidence of differential teacher effects. Therefore, all estimates reported in tables 2–6 are estimates of the persistence of teacher effects. If all students, that is, lower- or higher-performing students, benefited equally from teacher effects, one would expect similar regression estimates across the achievement distribution.

Results of the quantile regression analysis are summarized in table 2. Specifically, table 2 reports regression estimates of teacher effects and their standard errors at the tenth, twenty-fifth, fiftieth, seventy-fifth, and ninetieth quantiles of the achievement distribution by grade. The last column of table 2 shows the sample sizes. Because the outcomes were standardized, the regression estimates indicate that changes in teacher effects correspond to changes in standard deviation units in achievement. All estimates were adjusted for covariate effects as indicated in equation (4). In first-grade mathematics, all regression estimates were positive and significantly different from zero, suggesting that increases in teacher effects in kindergarten significantly increased student achievement in first grade. Overall, the estimates in the upper tail were larger in magnitude than the estimates in the lower tail of the achievement distribution. In second-grade mathematics, all regression estimates were also positive and significantly different from zero, suggesting that increases in teacher effects in first grade significantly increased student achievement in second grade. Again, overall, the estimates in the upper tail were larger in magnitude than the estimates in the lower tail of the achievement distribution.

TABLE 2

*Persistence of Teacher Effects Estimates in Mathematics and Reading at Various Quantiles: Linear Effects*

| | QUANTILE | | | | | |
|---|---|---|---|---|---|---|
| GRADE | Tenth | Twenty-Fifth | Fiftieth | Seventy-Fifth | Ninetieth | $N$ |
| Mathematics: | | | | | | |
| 1 | .237* | .374* | .491* | .443* | .452* | 4,358 |
| | (.076) | (.070) | (.090) | (.086) | (.135) | |
| 2 | .551* | .573* | .535* | .594* | .844* | 4,638 |
| | (.118) | (.104) | (.110) | (.148) | (.175) | |
| 3 | .432* | .505* | .509* | .642* | .706* | 4,780 |
| | (.118) | (.134) | (.107) | (.108) | (.126) | |
| 4 | .801* | .679* | .505* | .438* | .556* | 4,215 |
| | (.157) | (.092) | (.074) | (.088) | (.093) | |
| Reading: | | | | | | |
| 1 | .228* | .360* | .604* | .527* | .300* | 4,255 |
| | (.069) | (.079) | (.102) | (.099) | (.091) | |
| 2 | .856* | .847* | .863* | .845* | .980* | 4,641 |
| | (.140) | (.156) | (.129) | (.121) | (.123) | |
| 3 | .774* | .842* | .841* | .712* | .846* | 4,797 |
| | (.150) | (.139) | (.113) | (.120) | (.122) | |
| 4 | 1.010* | .803* | .810* | .738* | .722* | 4,134 |
| | (.180) | (.138) | (.099) | (.065) | (.108) | |

NOTE.—Standard errors of estimates are in parentheses.
* $p < .05$.

Similarly, in third-grade mathematics, all regression estimates were positive and significantly different from zero, suggesting that increases in teacher effects in second grade increased student achievement in third grade. The estimates in the upper tail were again larger in magnitude than the estimates in the lower tail of the achievement distribution. In fourth-grade mathematics, all regression estimates were positive and significantly different from zero, suggesting that increases in teacher effects in third grade significantly increased student achievement in fourth grade. However, overall, the estimates in the lower tail were larger in magnitude than the estimates in the upper tail of the achievement distribution. In sum, across all grades, students benefited from teacher effects. It appears that in the second, third, and fourth grades, the teacher effects were more pronounced than in the first grade, with the majority of estimates suggesting achievement increases larger than one-half of a standard deviation. Values of indexes of goodness of fit such as the pseudo *R*-squared ranged between 7 and 10 percent across quantiles and grades.

The lower panel of table 2 reports the results for reading achievement. In first-grade reading, all regression estimates were positive and significantly different from zero, suggesting that increases in teacher effects in kindergarten

TABLE 3

*Persistence of Residualized Teacher Effects Estimates in Mathematics and Reading and Various Quantiles: Linear Effects*

| | QUANTILE | | | | | |
|---|---|---|---|---|---|---|
| GRADE | Tenth | Twenty-Fifth | Fiftieth | Seventy-Fifth | Ninetieth | $N$ |
| Mathematics: | | | | | | |
| 2 | .530* | .572* | .471* | .526* | .562* | 3,327 |
| | (.133) | (.120) | (.111) | (.159) | (.231) | |
| 3 | .272* | .267* | .268* | .322* | .459* | 3,806 |
| | (.111) | (.095) | (.107) | (.072) | (.101) | |
| 4 | .590* | .437* | .343* | .250* | .312* | 3,603 |
| | (.112) | (.076) | (.060) | (.082) | (.117) | |
| Reading: | | | | | | |
| 2 | .755* | .724* | .656* | .657* | .862* | 3,331 |
| | (.126) | (.106) | (.107) | (.130) | (.193) | |
| 3 | .584* | .549* | .542* | .466* | .703* | 3,821 |
| | (.131) | (.161) | (.147) | (.112) | (.189) | |
| 4 | .772* | .530* | .483* | .470* | .315* | 3,587 |
| | (.149) | (.135) | (.117) | (.088) | (.126) | |

NOTE.—Standard errors of estimates are in parentheses.
* $p < .05$.

increased student achievement in first grade significantly. The patterns were similar in the second, third, and fourth grades. All regression estimates across grades were positive and statistically significant. The estimates were overall larger than those in mathematics, which indicates stronger associations between teacher effects and achievement in reading than in mathematics. In the first, second, and third grades, the teacher effects seemed to be uniform across different quantiles of the achievement distribution; in the fourth grade, it appeared that lower-performing students may have benefited more from having effective teachers in the third grade than other students. Across all grades, students benefited from teacher effects, and it appears that in the second, third, and fourth grades, the teacher effects were more pronounced than in the first grade, with the majority of estimates suggesting achievement increases larger than three-fourths of a standard deviation. Values of indexes of goodness of fit such as the pseudo *R*-squared ranged between 9 and 14 percent across quantiles and grades.

The estimates of the persistence of residualized teacher effects are summarized in table 3. By and large the results are qualitatively similar to those reported in table 2. All estimates were positive and significant but smaller in magnitude than those in table 2. In fourth-grade mathematics and reading, the estimates were more pronounced and nearly twice as large in the tenth quantile than in the ninetieth quantile. Values of indexes of goodness of fit

TABLE 4

*Persistence of Teacher Effects Estimates in Mathematics and Reading at Various Quantiles: Quadratic Effects*

| | QUANTILE | | | | | |
|---|---|---|---|---|---|---|
| GRADE | Tenth | Twenty-Fifth | Fiftieth | Seventy-Fifth | Ninetieth | $\mathcal{N}$ |
| Mathematics: | | | | | | |
| 1 | −.209 | .063 | .197 | .462 | .616 | 4,358 |
| | (.290) | (.367) | (.347) | (.353) | (.336) | |
| 2 | −.063 | −.237 | −.008 | −.070 | −.384 | 4,638 |
| | (.217) | (.229) | (.231) | (.266) | (.385) | |
| 3 | −.073 | −.130 | .092 | −.290 | −.145 | 4,780 |
| | (.241) | (.254) | (.371) | (.408) | (.447) | |
| 4 | −.336 | −.532* | −.368 | −.362 | −.078 | 4,215 |
| | (.265) | (.223) | (.220) | (.250) | (.240) | |
| Reading: | | | | | | |
| 1 | −.270 | −.294 | −.503 | −.063 | −.120 | 4,255 |
| | (.157) | (.154) | (.270) | (.316) | (.295) | |
| 2 | −.049 | .011 | −.221 | .274 | .189 | 4,641 |
| | (.343) | (.447) | (.488) | (.722) | (.772) | |
| 3 | −.758 | −.560 | −.599 | −.757 | −1.000 | 4,797 |
| | (.390) | (.446) | (.387) | (.503) | (.598) | |
| 4 | −1.217 | −.972 | −.652 | −.423 | −.781 | 4,134 |
| | (.744) | (.529) | (.483) | (.442) | (.516) | |

NOTE.—Standard errors of estimates are in parentheses.
* $p < .05$.

such as the pseudo *R*-squared ranged between 6 and 8 percent in mathematics and between 8 and 12 percent in reading.

## Nonlinear Teacher Effects

We also examined possible nonlinear teacher effects by including quadratic terms for teacher effects in equation (4). These estimates were also adjusted for covariate effects. The results of this analysis are summarized in table 4, which has the same structure as tables 2 and 3. The linear effects estimates are not included in the table for simplicity, but they were all positive and significant.

In first-grade mathematics, the quadratic estimates were typically positive but not significant at the .05 level. The estimates in the second, third, and fourth grades were negative and not significant. The only exception was the twenty-fifth quantile estimate at fourth grade, which was negative and significant at the .05 level, suggesting that in the low quartile of the distribution, student achievement increased when teacher effects increased but at a de-

TABLE 5

*Persistence of Residualized Teacher Effects Estimates in Mathematics and Reading at Various Quantiles: Quadratic Effects*

| | QUANTILE | | | | | |
|---|---|---|---|---|---|---|
| GRADE | Tenth | Twenty-Fifth | Fiftieth | Seventy-Fifth | Ninetieth | $N$ |
| Mathematics: | | | | | | |
| 2 | .345 | .116 | .215 | .091 | .059 | 3,327 |
| | (.221) | (.227) | (.259) | (.261) | (.312) | |
| 3 | −.084 | −.210 | −.155 | −.154 | −.412 | 3,806 |
| | (.287) | (.204) | (.249) | (.308) | (.383) | |
| 4 | −.161 | −.114 | −.161 | −.208 | −.189 | 3,603 |
| | (.216) | (.189) | (.097) | (.118) | (.246) | |
| Reading: | | | | | | |
| 2 | −.542 | .093 | −.081 | −.415 | −.753 | 3,331 |
| | (.591) | (.426) | (.515) | (.527) | (.583) | |
| 3 | −.092 | −.001 | −.124 | −.336 | −.619 | 3,821 |
| | (.558) | (.535) | (.497) | (.330) | (.641) | |
| 4 | −.396 | −.559 | −.363 | −.496 | −.398 | 3,587 |
| | (.737) | (.440) | (.487) | (.505) | (.628) | |

NOTE.—Standard errors of estimates are in parentheses.

creasing rate. The results for reading scores were qualitatively similar. The nonlinear estimates were negative and not significantly different from zero across grades. The estimates were larger in the third and fourth grades, indicating a higher likelihood of nonlinear effects in these grades in reading achievement. The quadratic estimates were consistently more pronounced in reading than in mathematics. Overall, these results provide very weak to no evidence of nonlinear teacher effects. Values of goodness of fit indexes were the same as those reported in table 2.

The quadratic estimates of the persistence of residualized teacher effects are summarized in table 5. By and large the results are qualitatively similar to those reported in table 4, only now none of the quadratic estimates are significant and the magnitude of these effects is much smaller than what was reported in table 4. Values of goodness of fit indexes were the same as those reported in table 3.

## Teacher Effects across Grades

Overall from kindergarten to fourth grade, nearly 10 percent of students consistently had high- or low-effective teachers (i.e., teachers in the top or bottom half of the distribution). The results of this analysis are reported in table 6. The results for fourth-grade mathematics scores are reported in the

TABLE 6

*Estimates of Persistence of Teacher Effects in Successive Grades: Kindergarten to Third Grade*

| GRADE 4 | QUANTILE | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | Tenth | Twenty-Fifth | Fiftieth | Seventy-Fifth | Ninetieth | $N$ |
| Mathematics: | | | | | | |
| K–3 teacher effect: top half | .432* | .323* | .221* | .286* | .314* | 2,297 |
| | (.071) | (.069) | (.052) | (.062) | (.078) | |
| K–3 teacher effect: bottom half | −.305* | −.278* | −.318* | −.350* | −.346* | 2,297 |
| | (.163) | (.058) | (.076) | (.043) | (.102) | |
| Reading: | | | | | | |
| K–3 teacher effect: top half | .317* | .285* | .255* | .183* | .271* | 2,259 |
| | (.073) | (.057) | (.051) | (.066) | (.083) | |
| K–3 teacher effect: bottom half | −.232 | −.226* | −.236* | −.187* | −.113 | 2,259 |
| | (.124) | (.064) | (.090) | (.079) | (.114) | |

NOTE.—Standard errors of estimates are in parentheses.
\* $p < .05$.

upper panel and for reading scores in the lower panel. The estimates are mean differences in standard deviation units between students who consistently had high- or low-effective teachers from kindergarten to third grade and students who did not. As expected, the estimates are positive for students who have had high-effective teachers and negative for those who have had low-effective teachers. The estimates indicated a significant advantage for students who have had high-effective teachers in successive grades both in mathematics and in reading. The advantage seemed larger for lower-performing students both in mathematics and in reading. However, the differences in the estimates at the lower and upper quantiles did not reach statistical significance. The benefit seemed larger in mathematics than in reading in the tails.

The students who have had low-effective teachers in successive grades were at a disadvantage, however. In particular, the disadvantage was significant for all students and was larger than one-fourth of a standard deviation in mathematics. The disadvantage was more pronounced in the tails of the mathematics distribution, which is alarming for lower-performing students in particular. In reading, the estimates were smaller and insignificant at the ninetieth quantile. That is, the disadvantage was less pronounced in reading especially for very high-performing students. Still these effects are overall not trivial both in mathematics and in reading and suggest that having high-effective teachers successively in early grades is beneficial, whereas having low-effective teachers successively in early grades is a disadvantage especially in mathematics. These results are important for lower-performing students in particular because these students need the additional boost from high-effective teachers the most. Values of indexes of goodness of fit such as the pseudo *R*-squared ranged between 5 and 7 percent.

Discussion

We investigated the persistence of teacher effects from grade to grade across the achievement distribution. We were interested in whether teacher effects are differential or uniform for lower-performing, medium, and higher-performing students. We used high-quality data from a four-year randomized experiment in which teachers and students were assigned randomly to classrooms within schools. The results of the analyses suggest that overall, in early grades, teacher effects in one grade lead to higher academic achievement in the following grade for lower-performing, medium, and higher-performing students. This finding supports the notion that teachers can increase achievement significantly for all students. In addition, teacher effects were not trivial and typically showed that students who had effective teachers in one year demonstrated a significant increase in their achievement the following year. The achievement gain was more pronounced in reading and reached three-fourths of a standard deviation in some grades.

There was not consistent evidence of differential teacher effects on student achievement, however. Overall, teacher effects seemed uniform across the achievement distribution, which suggests that students at different achievement levels benefited equally from teachers. There were some exceptions, however. In fourth grade the estimates were larger for lower-performing students. The teacher effect benefit in first-grade mathematics, however, indicated an advantage for higher-performing students (e.g., ninetieth quantile) and was nearly twice as large as that at the tenth quantile. These estimates were not different in a statistical sense. The estimates were also larger for higher-performing students in the second and third grades, but the differences in these estimates were not significant. In sum, it appears that all students benefited from teachers similarly and significantly.

We also explored whether the effects of teachers were nonlinear by adding both linear and quadratic terms in the quantile regression models. These results showed no systematic evidence of nonlinear effects. The overwhelming majority of the nonlinear estimates were not significant and suggested that teacher effects are linear. Notice that in all quantile regression models we included the inverse Mills ratio as a covariate to adjust for possible selection. Virtually in all models the selection coefficient was positive and in many instances significant, indicating some positive selection of individuals from grade to grade. Thus, it appears that attrition may have resulted in some positive selection.

Teacher effects both linear and quadratic seem consistently larger in reading than in mathematics. This is an interesting finding given that the students in the same classroom are taught mathematics and reading by the same teacher. Student selection is also unlikely since virtually the same samples of students

took the SAT-9 mathematics and reading tests. One explanation may be that teachers typically put more emphasis on reading than on mathematics in early grades and that the pedagogy of reading is heavily infused in early grades. Familiarity with the basic mechanisms of reading, vocabulary growth, and systematic practice in reading take place in early grades. In addition, basic reading skills such as decoding are developed in early grades and lay the foundation for later, more advanced reading skills such as comprehension. A related point is that teachers who teach in early grades may be better prepared to teach reading than mathematics. Also, schools may stress the importance of focusing more on reading in early grades. Unfortunately, classroom observations or teacher logs were not available in Project STAR, and therefore, it is impossible to know the actual teaching practices that took place in each classroom. Regardless of the mechanism, the empirical evidence points to larger persistence of teacher effects in reading.

The results from the longitudinal analysis are also informative. Students who have had high-effective teachers in successive grades benefited at least one-fourth of a standard deviation in fourth-grade mathematics. The advantage seems larger for lower-performing students. The students who have had low-effective teachers in successive grades, however, were at a disadvantage that was larger than one-fourth of a standard deviation in the tails of the fourth-grade mathematics distribution. This is especially concerning for lower-performing students. The disadvantage was less pronounced in reading, however, especially for very high-performing students. Generally, these findings suggest that having high-effective teachers successively in early grades is beneficial, whereas having low-effective teachers successively in early grades could be potentially harmful in mathematics, especially for lower-performing students. The findings stress the importance of assigning effective teachers to classrooms with higher proportions of lower-performing students.

Unfortunately, we could not control for teacher effects in the fourth grade. There was no information about teachers in the fourth-grade data, and thus, we were able to control only for student characteristics and school urbanicity and composition (e.g., percentage of minority or disadvantaged students). For grades 1–3, however, we were able to include in the models covariates such as teacher race, experience, and education, and therefore, we adjusted the persistence of teacher effects estimates. Still, we were not able to control for all current-grade teacher effects since the information about teacher characteristics was limited. As a result, covariates that could have affected our estimates such as teacher peer effects, teacher turnover, and teacher tenure in the same grade level were not controlled for in this analysis. This is a potential limitation of the study. Further, in all models we controlled for school urbanicity and school composition such as percentage of minority and disadvantaged students in the school. These were the only school variables avail-

able. However, other unobserved school variables such as percentage of effective or high-quality teachers in the school could have affected our estimates. This is also a potential limitation of the study. Finally, another potential limitation is that the teacher effects were computed assuming no measurement error, which is an assumption that may not hold exactly and could affect our regression coefficients in the quantile regression analysis.

Teacher effects were estimated assuming a constant variance for the entire achievement distribution. This assumption may be restrictive in that the constant variance may represent well the data around the middle of the distribution but not necessarily other data, especially in the tails. If this assumption does not hold for the tails of the distribution and the variance in the tails is larger, then our variance estimate of the teacher effects may be conservative. As a result, it is possible that the prediction (i.e., teacher effects in one year predicting student achievement the following year) is underestimated because of the restriction of range in the predictor. Still our empirical estimates from the quantile regression are significant and not trivial in magnitude.

In addition, one way to check whether the variance of the teacher effects is constant across the achievement distribution in one grade is to treat initial achievement (i.e., pretreatment scores) as a random effect at the teacher level. Essentially, this model would assume that initial achievement varies across teachers and therefore interacts with teachers. A significant variance of this interaction random effect would suggest that achievement is not consistent across teachers. Unfortunately, we were unable to investigate this because pretest scores were not available in kindergarten. In addition, even if pretest scores were available, it is unclear that they would vary across teachers within schools given random assignment of students to classes within schools.

To conclude, although this study demonstrates that the persistence of teacher effects has a positive impact on student achievement for all students, there is not much evidence of differential teacher effects except for the fourth grade. It does not seem that lower-performing students benefit more from having effective teachers in the previous grade than other students. However, lower-performing students benefit from teachers at least as much as other students, which is promising. The longitudinal analysis revealed a larger detriment in reading for lower-performing students who have had low-effective teachers in successive grades. The present study does not unravel the mechanism through which teacher effects persist and affect student achievement. This is partly due to the general definition of teacher effects, which does not allow examination of associations between observed teacher characteristics or teaching practices and student achievement. Data about teaching practices in classrooms are unfortunately not available. Such data could have helped identify the mechanism of teacher effectiveness because they typically include information about instructional processes and interactions among students and

between students and teachers. A well-designed study with the objective of collecting high-quality micro-level data in the classroom would provide invaluable information about the mechanism of teacher effectiveness.

## References

Ammermueller, Andreas, and Jörn-Steffen Pischke. 2009. "Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study." *Journal of Labor Economics* 27 (3): 315–48.

Ballou, Dale, William Sanders, and Paul Wright. 2004. "Controlling for Student Background in Value-Added Assessment of Teachers." *Journal of Educational and Behavioral Statistics* 29 (1): 37–65.

Borman, Geoffrey D., and Steven M. Kimball. 2005. "Teacher Quality and Educational Equality: Do Teachers with Higher Standards-Based Evaluation Ratings Close Student Achievement Gaps?" *Elementary School Journal* 106 (1): 3–20.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, Jonah Rockoff, and James Wyckoff. 2008. "The Narrowing Gap in New York City Teacher Qualifications and Its Implications for Student Achievement in High-Poverty Schools." *Journal of Policy Analysis and Management* 27 (4): 793–818.

Braun, Henry I. 2005. "Value-Added Modeling: What Does Due Diligence Require?" In *Value-Added Models in Education: Theory and Applications*, ed. Robert W. Lissitz. Maple Grove, MN: JAM Press.

Bryk, Anthony S., and Stephan W. Raudenbush. 1988. "Toward a More Appropriate Conceptualization of Research on School Effects: A Three-Level Hierarchical Linear Model." *American Journal of Education* 97 (November): 65–108.

Buchinsky, Moshe. 1998. "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research." *Journal of Human Resources* 33 (1): 89–126.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vidgor. 2006. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." *Journal of Human Resources* 41 (4): 778–820.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. 1966. *Equality of Educational Opportunity*. Washington, DC: US Government Printing Office.

Cronbach, Lee J., and Richard E. Snow. 1977. *Aptitudes and Instructional Methods*. New York: Irvington.

Freedman, David A. 2006. "Statistical Models for Causation: What Inferential Leverage Do They Provide?" *Evaluation Review* 30 (6): 691–713.

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

Goldhaber, Dan D., and Dominic J. Brewer. 1997. "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity." *Journal of Human Resources* 32 (3): 505–23.

Good, Thomas L. 1979. "Teacher Effectiveness in the Elementary School." *Journal of Teacher Education* 30 (2): 52–64.

Good, Thomas, and Jere E. Brophy. 1987. *Looking in Classrooms*. New York: Harper & Row.

Greenwald, Rob, Larry V. Hedges, and Richard D. Laine. 1996. "The Effect of School Resources on Student Achievement." *Review of Educational Research* 66 (3): 361–96.

Hanushek, Eric A. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24 (3): 1141–77.

Hanushek, Eric A. 1992. "The Trade-off between Child Quantity and Quality." *Journal of Political Economy* 100 (1): 84–117.

Hanushek, Eric A. 1999. "Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects." *Educational Evaluation and Policy Analysis* 21 (2): 143–63.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. 2003. "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics* 18 (5): 527–44.

Hanushek, Eric A., and Steven G. Rivkin. 2010. "Generalizations about Using Value-Added Measures of Teacher Quality." *American Economic Review Papers and Proceedings* 100 (2): 267–71.

Hao, Lingxin, and Daniel Q. Naiman. 2007. *Quantile Regression*. Thousand Oaks, CA: Sage.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–61.

Hill, Heather C., Brian Rowan, and Deborah Loewenberg Ball. 2005. "Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement." *American Educational Research Journal* 42 (2): 371–406.

Kennedy, Mary M. 2008. "Contributions of Qualitative Research to Research in Teacher Qualifications." *Educational Evaluation and Policy Analysis* 30 (4): 344–67.

Koenker, Roger, and Gilbert Bassett Jr. 1978. "Regression Quantiles." *Econometrica* 46 (3): 33–50.

Konstantopoulos, Spyros. 2009. "Effects of Teachers on Minority and Disadvantaged Students' Achievement in the Early Grades." *Elementary School Journal* 110 (1): 92–113.

Konstantopoulos, Spyros. 2011a. "How Consistent Are Class Size Effects?" *Evaluation Review* 35 (1): 71–92.

Konstantopoulos, Spyros. 2011b. "Teacher Effects in Early Grades? Evidence from a Randomized Experiment." *Teachers College Record* 113 (7): 1541–65.

Konstantopoulos, Spyros, and Vicki Chung. 2011. "The Persistence of Teacher Effects in Elementary Grades." *American Educational Research Journal* 48 (2): 361–86.

Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114 (2): 497–532.

Lefgren, Lars. 2004. "Educational Peer Effects and the Chicago Public Schools." *Journal of Urban Economics* 56 (2): 169–91.

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

Mashburn, Andrew J., Laura M. Justice, Jason T. Downer, and Robert C. Pianta. 2009. "Peer Effects on Children's Language Achievement during Pre-kindergarten." *Child Development* 80 (3): 686–702.

McCaffrey, Daniel F., J. R. Lockwood, Daniel Koretz, Thomas A. Louis, and Laura Hamilton. 2004. "Models for Value-Added Modeling of Teacher Effects." *Journal of Educational and Behavioral Statistics* 29 (1): 67–101.

Meyer, Robert H. 1997. "Value-Added Indicators of School Performance: A Primer." *Economics of Education Review* 16 (3): 283–301.

Murnane, Richard, and Barbara R. Phillips. 1981. "What Do Effective Teachers of Inner-City Children Have in Common?" *Social Science Research* 10 (1): 83–100.

Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos. 2000. "The Effects of Small Classes on Achievement: The Results of the Tennessee Class Size Experiment." *American Educational Research Journal* 37 (1): 123–51.

Nye, Barbara, Spyros Konstantopoulos, and Larry V. Hedges. 2004. "How Large Are Teacher Effects?" *Educational Evaluation and Policy Analysis* 26 (3): 237–57.

Raudenbush, Stephen W. 2004. "What Are Value-Added Models Estimating and What Does This Imply for Statistical Practice?" *Journal of Educational and Behavioral Statistics* 29 (1): 121–29.

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

Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics* 125 (1): 165–214.

Rowan, Brian, Richard Correnti, and Robert J. Miller. 2002. "What Large-Scale, Survey Research Tells Us about Teacher Effects on Student Achievement: Insights from the Prospects Study of Elementary Schools." *Teachers College Record* 104 (8): 1525–67.

Rubin, Donald B., Elizabeth A. Stuart, and Elaine L. Zanutto. 2004. "A Potential Outcomes View of Value-Added Assessment in Education." *Journal of Educational and Behavioral Statistics* 29 (1): 103–16.

Sanders, William L., and June C. Rivers. 1996. "Cumulative and Residual Effects of Teachers on Future Student Academic Achievement." Unpublished manuscript, Value-Added Research and Assessment Center, University of Tennessee, Knoxville.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64 (5): 416–24.

Weiss, Michael. 2010. "The Implications of Teacher Selection and the Teacher Effect in Individually Randomized Group Treatment Trials." *Journal of Research on Educational Effectiveness* 3 (4): 381–405.