DISCUSSION PAPER SERIES

IZA DP No. 5974

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

Spyros Konstantopoulos
Min Sun

September 2011

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

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

Spyros Konstantopoulos
Michigan State University

Min Sun
Virginia Tech

Discussion Paper No. 5974
September 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 5974
September 2011

ABSTRACT

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

We examined the persistence of teacher effects from grade to grade on lower-performing students using high-quality experimental data from Project STAR, where students and teachers were assigned randomly to classrooms of different sizes. The data included information about mathematics and reading scores and student demographics such as gender, race, and SES. Teacher effects were computed as residual classroom achievement within schools and within grades. Then, teacher effects were used as predictors of achievement in following grades and quantile regression was used to estimate their persistence. Results consistently indicated that all students benefited similarly from teachers. Overall, systematic differential teacher effects were not observed and it appears that lower-performing students benefit as much as other students from teachers. In fourth grade there was some evidence that lower-performing students benefit more from effective teachers. Results from longitudinal analyses suggested that having effective teachers in successive grades is beneficial to all students and to lower-performing students in particular in mathematics. However, having low-effective teachers in successive grades is detrimental to all students and to lower-performing students in particular in reading.

JEL Classification: I20

Keywords: teacher effects, low-achievers, quantile regression

Corresponding author:

Spyros Konstantopoulos
College of Education
Michigan State University
Erickson Hall
East Lansing, MI 48824
USA
E-mail: spyros@msu.edu

Since the Coleman Report much of educational research has focused on identifying school-related factors that affect student learning, and many of the school policy initiatives have attempted to ensure that school resources are allocated adequately to schools. One factor that is 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. Because the U.S. educational system is also designed to provide equal access to school resources to all students 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) was 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 teachers, their effectiveness in particular. NCLB has mandated state plans to improve teacher effectiveness, with the underlying belief that effective teachers can improve achievement especially for lower-performing students. It is a timely then, to examine how teachers affect lower-performing students and whether these types of students benefit more from teachers.

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

Findings about the 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 & Philips, 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 & 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 documented well. In this study, we examined the persistence of the effects teachers have on lower-performing students from grade to grade using high-quality experimental data from Project STAR (Student Teacher Achievement Ratio) (Krueger, 1999; Nye, Hedges, & Konstantopoulos, 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 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 related notion 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 upon 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. 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 will suggest whether the persistence of teacher effects is considerable to be of policy relevance. Second, knowing whether all or some students benefit similarly or differently from having had effective teachers is potentially valuable. This information will provide an explicit inference about the generality of the results and will point 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 of research 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. Economics have demonstrated a positive association between teacher experience and student achievement (Clotfelter, Ladd, & Vidgor, 2006). Education researchers have examined the effect of teacher content knowledge on student achievement (Hill, Rowan, & Ball, 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 SES (socioeconomic status) and teacher characteristics such as experience. Kennedy (2008) also found that teacher content knowledge seems to benefit students. Finally, researchers have also provided evidence that National Board certified teachers seem to be more effective than other teachers (Goldhaber & 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 & 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 & Brophy, 1987). Reviewers of teacher effects from process-product studies have concluded that effective teachers influenced academic achievement for all students substantially (Good, 1979). In addition, other studies have documented that teachers with higher evaluation scores in their teaching also had higher classroom achievement means and contributed in closing the achievement gap between lower and higher SES students in some grades (Borman & Kimball, 2005). Improvements in teacher qualifications also seem to increase student achievement especially in poor schools (Boyd, Lankford, Loeb, Rockoff, & Wyckoff, 2008)

The third area is known as value-added research. Value-added models have gained considerable attention the last 15 years mainly because of the urgency to use achievement scores to determine teacher effects on student outcomes, and especially with the passing 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 that 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 post-test 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 over adjust the teacher effects estimates (Ballou, Sanders, & Wright, 2004; Sanders & 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 don't necessarily eliminate all possible confounding effects completely because unobservables may still be related with teacher effects (Braun, 2005). Rubin, Stuart and Zanutto (2004) provided a thoughtful discussion about causal inferences of teacher and school effects. In their discussion Rubin et al. 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, but should more likely be conceptualized as descriptive measures of teacher effects. Along those lines, 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, Sanders, & Wright, 2004; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; Sanders & 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 & Chung, 2011).

Studies of the 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 & Brewer, 1997; Murnane & Phillips, 1981; Rowan, Correnti, & Miller, 2002). A recent study provided also 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 approximately from one-tenth to one-fifth of a standard deviation and in mathematics from one-tenth to one-third of a standard deviation (Hanushek & Rivkin, 2010).

One “caveat” however of 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 a small proportion of the variation in teacher effectiveness (Konstantopoulos, 2011a; Rivkin et al., 2005). For example, Konstantopoulos found that teacher education and experience explained less than one percent of the variation in teacher effects in early grades. These findings suggest that the majority of the variability in teacher effects remains unobserved and it is not captured by observed teacher characteristics. It is possible that the teacher characteristics typically measured are easy 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 is so poor that the effects are attenuated.

Limitations of Previous Work

Students are frequently assigned to teachers based on their characteristics such as achievement. In turn, 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 compensatory strategy. This non-random 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 advantaged students are more likely to have highly qualified teachers than other studies, which biases the association between teacher characteristics and achievement.

Because of the 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 on student achievement, even important covariates such as prior achievement and SES do not completely eliminate differences in all 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 can 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 would not be systematic. Random assignment of teachers to classes is also important and 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 satisfies 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 depend 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 or students. Unfortunately, no pretest scores were collected in Project STAR, so it was not possible to examine differences in pre-kindergarten 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. Krueger 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, 2011b).

Even if we assume that random assignment across classroom types were 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 than those who dropped out of regular type classes (Kruger, 1999). In longitudinal designs such as Project STAR one way to measure the effects of differential attrition is by imputing 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 with school achievement and school composition (i.e., proportion of minority or disadvantaged students) (Konstantopoulos, 2011b). In this paper we attempted to adjust for possible selection from one grade to the next using the Heckman method as illustrated in the analysis section (Heckman, 1979).

Method

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 teacher and teacher aids, was about $12 million. The state of Tennessee paid for hiring 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 kindergarteners in more than 300 classrooms in 79 elementary schools in 42 districts in Tennessee participated. The sample included a broad range of schools and districts (e.g., urban, rural, wealthy, and poor). Districts had to agree to participate for four years, allow school visits for verification of class sizes, interviewing, 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 to 17 students), regular classes (with 22 to 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 grades 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.

Analysis

Computing Teacher Effects in Each Grade

The main objective of the 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 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. Because of 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. 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 where 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 (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 (e.g., 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 via variables that represent peer effects. We included in our model peer effects for variables such as gender, race, SES, and age (see specification in equation 1 below). A typical way of modeling peer effects is by computing aggregate measures of student variables for each classroom within each school for all students in the classroom except a specific student (see Amemueler & Pischke, 2009; Mashburn, Justice, Downer, & Pianta, 2009). To illustrate the process suppose that there are 21 students in a classroom. To compute the peer effect index for the 21st 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 lunch and then divided that number by 20. If 10 of the 20 students were eligible for free or reduced price lunch in the classroom then the peer effect for the 21st 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 in statistical models that measure teacher of schools effects as covariates. It is difficult to know whether that proportion of the between-teacher variance explained the student variables is solely attributed 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 and that our assumption that the between-teacher variance explained by student variables is solely due 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 grade. In mathematics the proportion of variance explained is overall very small 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 employing a three-level model (Bryk & 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 (i.e., kindergarten, first, second, or third grade) separately. Hence, for each grade the one-level regression equation for student *i*, in class *j*, in school *k* is

$$\begin{aligned}Y_{ijk} &= \beta_{000} + \beta_{100}FEMALE_{ijk} + \beta_{200}LOWSES_{ijk} + \beta_{300}MINORITY_{ijk} + \beta_{400}AGE_{ijk} + \\ &\beta_{500}PEERFEM_{ijk} + \beta_{600}PEERLOWSES_{ijk} + \beta_{700}PEERMINORITY_{ijk} + \beta_{800}PEERAGE_{ijk} + \\ &\beta_{010}SMALL_{jk} + \beta_{020}AIDE_{jk} + \varepsilon_{ijk} + \xi_{0jk} + \eta_{00k}\end{aligned} \quad (1)$$

where Y_{ijk} represents student achievement in mathematics or reading, *FEMALE* is a dummy variable for gender (i.e., female), *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, ε_{ijk} is a student-specific random effect, ξ_{0jk} is a classroom-specific random effect, and η_{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, ξ, 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 equation 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 (i.e., the current grade aggregate classroom achievement) is likely endogenous and is differentiated from the aggregate classroom measures of family background (Hanushek, Kain, Markman, & Rivkin, 2003). Specifically, the student and peer achievement 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 equation 1) that do not include current grade aggregate classroom achievement (Amemueler & 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

$$
\begin{aligned}
&Y_{ijk} = \beta_{000} + \beta_{100} FEMALE_{ijk} + \beta_{200} LOWSES_{ijk} + \beta_{300} MINORITY_{ijk} + \\
&\beta_{400} AGE_{ijk} + \beta_{500} PRACHIEVEMENT + \beta_{600} PEERFEM_{ijk} + \\
&\beta_{700} PEERLOWSES_{ijk} + \beta_{800} PEERMINORITY_{ijk} + \beta_{900} PEERAGE_{ijk} \\
&+\beta_{100} PEERPRACHIEVEMENT + \beta_{010} SMALL_{jk} + \beta_{020} AIDE_{jk} + \widehat{\varepsilon}_{ijk} + \widehat{\xi}_{0jk} + \widehat{\eta}_{00k} \qquad (2)
\end{aligned}
$$

where *PRACHIEVEMENT* indicates previous achievement (e.g., mathematics) and *PEERACHIEVEMENT* 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

Once the teacher effects (i.e., $\xi's$) were computed for kindergarten, first, second, and third grades they were then used as predictors of student achievement in subsequent grades (i.e., 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 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 that those who left the experiment, we tried to adjust for potential selection. One way to control for selection directly in a regression model is via the Heckman model (Heckman, 1979). The 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

$$probit\left(\pi\right) = \beta_0 + \mathbf{XB} \tag{3}$$

where β_0 is the constant, **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 Mill's ratio or lambda (λ), which we included as a covariate in the second stage achievement regressions to adjust for possible non-random selection of students. 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 than those in the probit model (i.e., 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 (e.g., 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 one through four (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 paper, it is possible that teachers in one grade have differential effects on average, lower, and higher-performing students in te 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 & 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 & 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 produce additional benefits in achievement for lower-performing students in grades one through four. 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, 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 (e.g., one through four). In each grade, mathematics and reading scores were regressed on teacher effects and other covariates. For example, in 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

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

where y is mathematics or reading scores, β_0 is a constant, TE is the teacher effect in previous grade, λ represents the sample selection, **ST** is a row vector of student characteristics such as gender, race/ethnicity, low SES, **CL** is a row vector of classroom/teacher characteristics such as type of classroom (small or regular with aide[1]), teacher race, education, and experience, **SC** is a row vector of school urbanicity indicators and school composition, (i.e., percent minority or disadvantaged students), and ε is the error. The betas are the regression coefficients that need to be estimated. The estimate of the teacher effect (β_1) is the most important for this study. For grades one through three 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 (e.g., 10^{th} and 25^{th} quantile), the middle (50^{th} quantile), and the upper tail (e.g., 75^{th} and top 90^{th} quantile) of the achievement distribution. Because our data have a nested structure, since students are nested within classrooms and schools, 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 (i.e., non-constant variation). Finally, 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, Sanders, & Wright, 2004; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; Sanders & 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. Sander and Rivers found that teacher effects were cumulative. In the same vein, we also 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 high 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 effective teachers as those who were in the top half of the teacher effects distribution in each grade (e.g., kindergarten, first, second, and third grade). Low effective teachers were defined as those who were in the bottom half of the teacher effects distribution in each grade (e.g., kindergarten through third grade). We coded the cumulative effects of low or high effective teachers as binary indicators that took the value of one if a student had low (or high) effective teachers in all four years (e.g., kindergarten through third grade) and zero otherwise. Then, we used quantile regression and regressed fourth grade mathematics or reading achievement on the cumulative teacher effects and other covariates (e.g., gender, race, SES, and class size in grade 4). We computed estimates of teacher effects in the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} quantiles. We coded teacher cumulative teacher effects using the top half or bottom half coding scheme, because only a small proportion of students received very low or very high effective teachers in consecutive grades. The quantile regressions were

$$Y_i = \beta_0 + \beta_1 TOP50_i + \beta_2 CLSIZE_i + \mathbf{ST}_i\mathbf{B}_3 + \mathbf{SC}_i\mathbf{B}_5 + \varepsilon_i,$$

or (5)

$$Y_i = \widetilde{\beta}_0 + \widetilde{\beta}_1 BOTTOM50_i + \widetilde{\beta}_2 CLSIZE_i + \mathbf{ST}_i\widetilde{\mathbf{B}}_3 + \mathbf{SC}_i\widetilde{\mathbf{B}}_5 + \widetilde{\varepsilon}_i,$$

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 most important coefficients in this analysis were β_1 and $\widetilde{\beta}_1$. 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 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 to 3. About 80 percent of students had white teachers and 35 percent of students had teachers with graduate degrees in grades k to 2. The average teacher experience ranged between 9 and 14 years. Approximately 30 percent of students attended inner city or urban schools, while the majority of students attended suburban or rural schools. In the fourth grade, nearly fifty percent of the students in the sample were female, about 40 percent of the students were eligible for free or reduced lunch, and 20 percent of the students were minorities. Nearly 85 percent of students attended suburban or rural schools, while only about 15 percent of students attended inner city or urban schools. The outcomes of interest were mathematics and reading scores that 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 equation 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).

Insert Table 1 Here

Variance Decomposition

During the first stage of the analysis (see equations 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 full model with covariates. In mathematics in grades k to 3 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 grades k to 3 the variances at the second the third levels changed slightly.

In reading in grades k to 3 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 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 grades k to 3 the variances at the second the third levels changed slightly.

Linear Teacher Effects

The main objective of the study was to examine whether teacher effects persisted from grade to grade and whether they were distributed uniformly across the achievement distribution, or whether there was evidence of differential teacher effects. Therefore, all estimates reported in Tables 2 to 6 are estimates of the persistence of teacher effects. If all students (e.g., 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 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} 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 increased student achievement in first grade significantly. The estimates in the upper tail were overall 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 increased significantly student achievement in second grade. Again, the estimates in the upper tail were overall larger in magnitude than the estimates in the lower tail of the achievement distribution. 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 also positive and significantly different from zero, suggesting that increases in teacher effects in third grade increased significantly student achievement in fourth grade. However, now the estimates in the lower tail were overall larger in magnitude than the estimates in the upper tail of the achievement distribution. Overall, 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.

Insert Table 2 Here

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 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, while 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 to those in Table 2. In fourth grade mathematics and reading the estimates were more pronounced and nearly twice as large in the 10^{th} quantile than in the 90^{th} quantile. Values of indexes of goodness of fit such as the pseudo R-squared ranged between 6 and 8 percent in mathematics, and between 8 and 12 percent in reading.

Insert Table 3 Here

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 25th quantile estimate at the 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 decreasing 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.

Insert Table 4 Here

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.

Insert Table 5 Here

Teacher Effects Across Grades

Overall from kindergarten to fourth grade nearly 10 percent of students had high or low effective teachers (i.e., teachers in the top or bottom half of the distribution) consistently. The results of this analysis are reported in Table 6. The results for fourth grade mathematics scores are reported in the 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 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 90th 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 reading, and suggest that having high effective teachers successively in early grades is beneficial, whilst 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.

Insert Table 6 Here

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, medium, and higher-performing students. We used high-quality data from a 4-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, 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., 90^{th} quantile) and was nearly twice as large as that at the 10^{th} 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 for 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 weak or no evidence of nonlinear effects. The overwhelming majority of the nonlinear estimates was not significant, and suggested that teacher effects are linear. In all models we included the inverse Mill's ratio as a covariate to adjust for possible selection. Virtually in all models the selection coefficient was positive and in many instances significant, indicating 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, whilst 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 only able to control for student characteristics and school urbanicity and composition (e.g., percent minority or disadvantaged students). For grades 1 through 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 many current grade teacher effects, since the information about teacher characteristics was limited. As a result, covariates such as teacher peer effects, teacher turnover, and teacher tenure in the same grade level that could have affected our estimates 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 percent of minority and disadvantaged students in the school. These were the only school variables available. However, other unobserved school variables such as percent of effective or high quality teachers in the school could have affected our estimates. This is also a potential limitation of the study. 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 due to 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., pre-treatment 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 pre-test scores were not available in kindergarten. In addition, even if pre-test 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 impact 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 at the classroom data would provide invaluable information about the mechanism of teacher effectiveness.

References

Ammermueller, A., & Pischke, J. S. (2009). Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study. *Journal of Labor Economics, 27,* 315-348.

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, *29,* 37-65.

Borman, G. D., & Kimball, S. M. (2005). Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps? *Elementary School Journal, 106,* 3-20.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management, 27,* 793-818.

Braun, H. (2005).Value-added modeling: What does due diligence require? In R. Lissitz (Ed.)*, Value-added models in education: Theory and applications* (pp 19-38). Maple Grove, MN: JAM Press.

Bryk, A. S., & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. *American Journal of Education, 97,* 65-108.

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

Clotfelter, C. T., Ladd, H. F., & Vidgor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources, 41,* 778-820.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Government Printing Office.

Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods. New York: Irvington.

Freedman, D. A. (2006). Statistical models for causation. What inferential leverage do they provide? *Evaluation Review*, *30,* 691-713.

Goldhaber, D. D., & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources*, *32,* 505-523.

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.

Good, T. (1979). Teacher effectiveness in the elementary school. *Journal of Teacher Education, 30*, 52-64.

Good, T., & Brophy, J. (1987). *Looking in classrooms.* New York: Harper & Row.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66,* 361-396.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24,* 1141-1177.

Hanushek, E. A. (1992). The tradeoff between child quantity and quality: Some empirical evidence. *Journal of Political Economy, 100,* 84-117.

Hanushek, E. 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,* 143-163.

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

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. 2003. Does peer ability affect student achievement? *Journal of Applied Econometrics* 18:527–44

Hao, L., & Naiman, D. Q. (2007). *Quantile Regression.* Thousand Oakes, CA: Sage.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica 47,* 153-161.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*, 371–406.

Kennedy, M. M. (2008). Contributions of qualitative research to research in teacher qualifications. *Educational Evaluation and Policy Analysis, 30,* 344-367.

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica* 46: 33–50.

Konstantopoulos S. (2009). Effects of teachers on minority and disadvantaged students' achievement in the early grades. *Elementary School Journal, 110* (1), 92-113.

Konstantopoulos, S. (2011a). Teacher effects in early grades? Evidence from a randomized experiment. *Teachers College Record, 113,* 1541-1565.

Konstantopoulos, S. (2011b). How consistent are class size effects? *Evaluation Review, 35,* 71 -92.

Konstantopoulos, S, & Chung, V. (2011). The persistence of teacher effects in elementary grades. *American Educational Research Journal, 48,* 361-386.

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics, 114,* 497-532.

Lefgren, L. (2004). Educational peer effects and the Chicago public schools. *Journal of Urban Economics, 56,* 169-191.

Manski, Charles F. 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60:531–42.

Mashburn, A. J., Justice, L. M., Downer, J. T., & Pianta, R. C. (2009). Peer Effects on Children's Language Achievement During Pre-Kindergarten. *Child Development, 80*, 686-702.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics, 29*, 67-101.

Meyer, R. H. 1997. Value-added indicators of school performance: A primer. *Economics of Education Review, 16,* 283-301.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (2000). The effects of small classes on achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37,* 123-151.

Nye, B., Konstantopoulos, S., & Hedges, L.V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26,* 237-257.

Murnane, R. J., & Phillips, B. R. (1981). What do effective teachers of inner-city children have in common? *Social Science Research, 10,* 83-100.

Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics, 29,* 121-129.

Rivkin, S. G., Hanushek, E. A., & Kain J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73,* 417-458.

Rothstein, J. (2010). Teacher quality in educational production: tracking, decay, and student achievement. *Quarterly Journal of Economics, 125,* 165-214.

Rowan, B., Correnti, R., & Miller, R. J. (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*, 1525-1567.

Rubin, D. B., Stuart, E. A, & Zanutto, E. L. 2004. A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics, 29*, 103-116.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* University of Tennessee Value-Added Research and Assessment Center.

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of Political Economy, 64,* 416-424.

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

Table 1. Descriptive Statistics of Variables of Interest Across Samples

	Grade				
Variable	K	1	2	3	4
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 = Socioeconomic Status

Table 2. Persistence of Teacher Effects Estimates in Mathematics and Reading at Various Quantiles: Linear Effects

	10th Quantile	25th Quantile	50th Quantile	75th Quantile	90th Quantile	N
Mathematics						
Grade:						
1	0.237* (0.076)	0.374* (0.070)	0.491* (0.090)	0.443* (0.086)	0.452* (0.135)	4358
2	0.551* (0.118)	0.573* (0.104)	0.535* (0.110)	0.594* (0.148)	0.844* (0.175)	4638
3	0.432* (0.118)	0.505* (0.134)	0.509* (0.107)	0.642* (0.108)	0.706* (0.126)	4780
4	0.801* (0.157)	0.679* (0.092)	0.505* (0.074)	0.438* (0.088)	0.556* (0.093)	4215
Reading						
Grade:						
1	0.228*(.0.069)	0.360* (0.079)	0.604* (0.102)	0.527* (0.099)	0.300* (0.091)	4255
2	0.856* (0.140)	0.847* (0.156)	0.863* (0.129)	0.845* (0.121)	0.980* (0.123)	4641
3	0.774* (0.150)	0.842* (0.139)	0.841* (0.113)	0.712* (0.120)	0.846* (0.122)	4797
4	1.010* (0.180)	0.803* (0.138)	0.810* (0.099)	0.738* (0.065)	0.722* (0.108)	4134

Note: Standard errors of estimates are in parenthesis

* $p < 0.05$

Table 3. Persistence of Residualized Teacher Effects Estimates in Mathematics and Reading at Various Quantiles: Linear Effects

	10th Quantile	25th Quantile	50th Quantile	75th Quantile	90th Quantile	N
Mathematics						
Grade:						
2	0.530* (0.133)	0.572* (0.120)	0.471* (0.111)	0.526* (0.159)	0.562* (0.231)	3327
3	0.272* (0.111)	0.267* (0.095)	0.268* (0.107)	0.322* (0.072)	0.459* (0.101)	3806
4	0.590* (0.112)	0.437* (0.076)	0.343* (0.060)	0.250* (0.082)	0.312* (0.117)	3603
Reading						
Grade:						
2	0.755* (0.126)	0.724* (0.106)	0.656* (0.107)	0.657* (0.130)	0.862* (0.193)	3331
3	0.584* (0.131)	0.549* (0.161)	0.542* (0.147)	0.466* (0.112)	0.703* (0.189)	3821
4	0.772* (0.149)	0.530* (0.135)	0.483* (0.117)	0.470* (0.088)	0.315* (0.126)	3587

Note: Standard errors of estimates are in parenthesis

* $p < 0.05$

Table 4. Persistence of Teacher Effects Estimates in Mathematics and Reading at Various Quantiles: Quadratic Effects

	10th Quantile	25th Quantile	50th Quantile	75th Quantile	90th Quantile	N
Mathematics						
Grade:						
1	-0.209 (0.290)	0.063 (0.367)	0.197 (0.347)	0.462 (0.353)	0.616 (0.336)	4358
2	-0.063 (0.217)	-0.237 (0.229)	-0.008 (0.231)	-0.070 (0.266)	-0.384 (0.385)	4638
3	-0.073 (0.241)	-0.130 (0.254)	0.092 (0.371)	-0.290 (0.408)	-0.145 (0.447)	4780
4	-0.336 (0.265)	-0.532* (0.223)	-0.368 (0.220)	-0.362 (0.250)	-0.078 (0.240)	4215
Reading						
Grade:						
1	-0.270 (0.157)	-0.294 (0.154)	-0.503 (0.270)	-0.063 (0.316)	-0.120 (0.295)	4255
2	-0.049 (0.343)	0.011 (0.447)	-0.221 (0.488)	0.274 (0.722)	0.189 (0.772)	4641
3	-0.758 (0.390)	-0.560 (0.446)	-0.599 (0.387)	-0.757 (0.503)	-1.000 (0.598)	4797
4	-1.217 (0.744)	-0.972 (0.529)	-0.652 (0.483)	-0.423 (0.442)	-0.781 (0.516)	4134

Note: Standard errors of estimates are in parenthesis

* p < 0.05

Table 5. Persistence of Residualized Teacher Effects Estimates in Mathematics and Reading at Various Quantiles: Quadratic Effects

	10th Quantile	25th Quantile	50th Quantile	75th Quantile	90th Quantile	N
Mathematics						
Grade:						
2	0.345 (0.221)	0.116(0.227)	0.215 (0.259)	0.091 (0.261)	0.059 (0.312)	3327
3	-0.084 (0.287)	-0.210 (0.204)	-0.155 (0.249)	-0.154 (0.308)	-0.412 (0.383)	3806
4	-0.161 (0.216)	-0.114 (0.189)	-0.161 (0.097)	-0.208 (0.118)	-0.189 (0.246)	3603
Reading						
Grade:						
2	-0.542 (0.591)	0.093 (0.426)	-0.081 (0.515)	-0.415 (0.527)	-0.753 (0.583)	3331
3	-0.092 (0.558)	-0.001 (0.535)	-0.124 (0.497)	-0.336 (0.330)	-0.619 (0.641)	3821
4	-0.396 (0.737)	-0.559 (0.440)	-0.363 (0.487)	-0.496 (0.505)	-0.398 (0.628)	3587

Note: Standard errors of estimates are in parenthesis

Table 6. Estimates of Persistence of Teacher Effects in Successive Grades (k to 3)

	10th Quantile	25th Quantile	50th Quantile	75th Quantile	90th Quantile	N
Grade 4 Mathematics						
K-3 Teacher Effect: Top 50	0.432* (0.071)	0.323* (0.069)	0.221* (0.052)	0.286* (0.062)	0.314* (0.078)	2297
K-3 Teacher Effect: Bottom 50	-0.305* (0.163)	-0.278* (0.058)	-0.318* (0.076)	-0.350* (0.043)	-0.346* (0.102)	2297
Grade 4 Reading						
K-3 Teacher Effect: Top 50	0.317* (0.073)	0.285* (0.057)	0.255* (0.051)	0.183* (0.066)	0.271* (0.083)	2259
K-3 Teacher Effect: Bottom 50	-0.232 (0.124)	-0.226* (0.064)	-0.236* (0.090)	-0.187* (0.079)	-0.113 (0.114)	2259

Note: Standard errors of estimates are in parenthesis

* $p < 0.05$