DISCUSSION PAPER SERIES

IZA DP No. 2893

How Long Do Teacher Effects Persist?

Spyros Konstantopoulos

June 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

How Long Do Teacher Effects Persist?

Spyros Konstantopoulos
Northwestern University
and IZA

Discussion Paper No. 2893
June 2007

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. 2893
June 2007

ABSTRACT

How Long Do Teacher Effects Persist?

Previous findings from experimental and non-experimental studies have demonstrated that teachers differ in their effectiveness. In addition, evidence from non-experimental studies has indicated that teacher effects can last up to five years. This study used high-quality data from a four-year randomized experiment in which teachers and students were randomly assigned to classes to examine whether teacher effects on student achievement persist over time. Teacher effects are defined as teacher specific residuals adjusted for student and treatment effects. Findings indicate that the teacher effects are cumulative and observed not only in the current or the following grade, but they endure up to three years in early elementary grades. The findings also suggest that teacher effects are important and their additive effects on student achievement are as large as the additive effects of small classes. Finally, teacher effects are larger in reading than in mathematics.

JEL Classification: I2

Keywords: teacher effects, experimental data, multi-level models

Corresponding author:

Spyros Konstantopoulos
School of Education and Social Policy
Northwestern University
2120 Campus Drive
Evanston, IL 60208
USA
E-mail: spyros@northwestern.edu

A fundamental goal of American education is to provide high quality educational experiences that facilitate academic growth for all students. Much of educational research has focused on identifying important school-related factors that affect student learning and many of the school policy initiatives have attempted to ensure that valuable school resources are allocated adequately across schools. What lies at the center of this line of research is the notion that school resources such as teachers *do matter* and can positively affect student achievement. Currently, with the passage of the No Child Left behind Act there is an urgency for test-based accountability, and test scores are widely used to hold teachers (and schools) accountable for student learning. An important task then, is to collect and examine evidence that accurately portrays the association between teacher effectiveness and student achievement.

One crucial question to educational research is whether teachers differ noticeably in their effectiveness as educators and pedagogues and in promoting their students' academic achievement. This question has been addressed in the teacher effects literature and recent experimental evidence has indicated that teachers *can indeed* make a difference (e.g., Nye, Konstantopoulos, & Hedges, 2004). However, this work has only assessed current teacher effects in specific years using cross-sectional samples of students.

Another important question to educational research is whether teacher effectiveness can affect student achievement over time. This question is related to the durability of teacher effectiveness on student achievement in subsequent years in school. One would expect that teacher effectiveness should have lasting benefits on student achievement because the influence that students receive over the years from their teachers is presumably cumulative in nature. If teaching has long-term meaningful and positive effects on student achievement then identification of more effective teachers as well as factors that cause teachers to be more effective is important both for educational research and reform. A corollary of the cumulative nature of teacher effectiveness is that not only the current, but the previous teachers matter as well. This is of critical importance for student learning especially in early school grades. If on the pother hand, teacher effectiveness has negligible long-term benefits on student achievement then perhaps only the current teacher effects make a different in student achievement in a specific year (assuming that there are significant teacher effects in a specific year).

Overall, the research evidence about teacher effects has been mixed. Some studies seem to suggest that teacher effects are negligible, while others suggest that they should be substantial (Greenwald, Hedges, & Laine, 1996; Hanushek, 1986). Recent experimental evidence has suggested that current teacher effects in specific years are large and meaningful in early grades (Nye et al., 2004). In addition, other non-experimental work has indicated that teacher effects are cumulative in nature in elementary school grades (e.g., Ballou, Sanders, & Wright, 2004; Sanders & Rivers, 1996). The present study examines the long-term benefits of teacher effects on student achievement in early grades using high quality experimental data from Project STAR (see Nye, Hedges, & Konstantopoulos, 2000). We are interested in determining the durability of teacher effects over time and whether teacher effects remain strong or fade over time. First, we briefly summarize some key results from previous research on teacher effects and indicate some of their limitations. Then, we discuss value-added models that examine teacher effects. Finally, we report some new experimental evidence, which is not subject to the same shortcomings as previous research.

Previous Research on Teacher Effects

There are two lines of research that have discussed teacher effects. The first tradition includes education production function studies that attempt to determine the relation of specific measured teacher characteristics (such as teacher experience, teacher education etc.) with student achievement. However, because parents choose neighborhoods in which to live, and hence their associated schools, according to tastes and resources (Tiebout, 1956), student and family backgrounds are confounded with naturally occurring teacher characteristics. Education production function studies (e.g., Coleman, et al., 1966), attempt to control for this confounding effect by using student and family background characteristics as covariates in regression models. A particularly important covariate is prior achievement, because it can be seen as summarizing the effects of individual background (including prior educational experiences) and family background up to that time. However, even this covariate may leave important characteristics of the student unmeasured. Some reviewers of the 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 there are positive effects of some of the resource characteristics such as teacher experience and teacher education (Greenwald, Hedges, and Laine, 1996).

The second tradition of research examines the *variation between classrooms* in achievement controlling for student background. These models typically use prior achievement as a covariate also, so they can be interpreted as measuring the variance in residualized student achievement gain across classrooms. That is, these variances represent the variation in achievement gain due to differences in teacher effectiveness. The underlying assumption is that the between-classroom variation in achievement is caused by teacher variation in effectiveness. Overall, the results of such studies have suggested that teacher effects are evident (see e.g., Goldhaber and Brewer, 1997; Murnane and Phillips, 1981; Rowan, Correnti, and Miller, 2002). In fact, a recent study provided experimental evidence about teacher effects defined in a similar manner that is consistent with those estimated in previous studies (Nye, Konstantopoulos, & Hedges, 2004). Notice however, that such models cannot identify specific teacher characteristics that are responsible for teacher effectiveness and define teacher effects as a general construct.

Nonetheless, in both lines of research it is difficult to interpret the estimates of teacher effects on achievement, even after controlling for previous achievement and student background (such as socioeconomic status) because the teacher effects may be confounded with unobserved individual, family, school, and neighborhood factors. It is not obvious that the observed covariates adequately control for all preexisting differences, including unobservable differences that are related to achievement such as motivation, among students assigned to different classrooms. This suggests that the teacher effects variable may be endogenous and its estimate may be biased. Students within schools are frequently assigned to teachers based on student characteristics such as achievement. In the same vein teachers are assigned to classrooms based on their characteristics such as experience. If for example more experienced teachers are assigned to classes composed of higher achieving students (e.g., as a privilege of seniority) or lower achieving students (e.g., as compensatory strategy of assigning human capital) the causal direction in the relation between teacher experience and student achievement is ambiguous.

The problems in interpretation of both research traditions discussed above would be eliminated if both students and teachers were randomly assigned to classes. Random assignment of students would assure that all observable and unobservable differences between students in different classes would be no larger than would be expected by chance. Random assignment of teachers to classes would assure that any differences in teacher characteristics are uncorrelated with classroom achievement (although this potential problem would also be substantially mitigated by the fact that randomization of students assures that there would be no large differences of student achievement across classrooms.) In this study we use data from the Tennessee class size experiment that satisfies both conditions of randomization.

Value Added Models

A main objective in teacher effects research is to identify the “best” way of estimating teacher effects on student achievement. A common way of estimating teacher effects is via value-added models. The underlying idea in value-added models is to examine the effects of teachers on students’ learning gains net of student effects (e.g., student background). That is such models intend to estimate the *unique* contribution or "value-added" of teachers on students' change or gain in learning. As Meyer (1997) argues the key objective in value-added research is to asses for example the 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 predictor variable represents teacher effects, and other variables such as pre-test measures of student achievement are included as covariates to adjust for previous knowledge or ability (Ballou, Sanders, & Wright, 2004; Sanders & Collins, 1996). These are often called the residualized change models. Other times the outcome is simply a change score between two time points (assuming the scales are equated) and including pre-test scores as a covariate is not necessary. A fundamental underlying assumption in value-added models is that the observed gains in student achievement are mainly due to teacher effects (e.g., instruction, teacher characteristics) and not to other student-related factors that are unobserved and hence not controlled for (see Schmidt, Houang, & McKnight, 2005). In principle value added models are hypothesized to provide more accurate estimates of teacher effectiveness. Value-added models have recently become more common because of the urgency to use achievement scores to determine teacher effectiveness on student outcomes.

Value-added models that examine the cumulative nature of teacher effects on student achievement have been used in previous work (e.g., Ballou et al., 2004; McCaffrey et al., 2004; Sanders & Collins, 1996). For example, Sanders and Collins used a value-added model to examine the predictive efficacy of teacher effects in grades 3, 4, and 5 on 5th grade achievement controlling for student achievement in grade 2. The authors found that the teacher effects were cumulative and important. In the present study we assess teacher effects over time following a model similar to that described by Sanders and Collins. For example, in some of our models we compute the predictive power of teacher effects in grades 1 and 2 on 3rd grade achievement controlling for student achievement in kindergarten. In our specifications, we also include gender, race, and SES effects and adjust for current teacher effects in grade 3. Although including pre-test scores as covariates, according to the value-added philosophy, is crucial, it should be noted that including pre-test scores should not be necessary in principle because we use experimental data from Project STAR. That is, in each grade students and teachers are randomly assigned to different types of classrooms within schools and hence, simply controlling for treatment effects and current teacher effects would most likely be sufficient (if randomization were successful). Two recent studies checked whether randomization was effective in Project STAR and found no evidence that randomization was not successful (Krueger, 1999; Nye, Kostantopoulos, & Hedges, 2004).

The Tennessee Class Size Experiment

The Tennessee class size experiment or Project STAR (Student-Teacher Achievement Ratio) is discussed in detail elsewhere (see, e.g., Nye, Hedges, and Konstantopoulos, 2000). The experiment involved students in 79 elementary schools in 42 school districts in Tennessee. Within each school, Kindergarten students were randomly assigned to classrooms in one of three treatment conditions: small classes (with

13 to 17 students), larger classes (with 22 to 26 students) or larger classes with a full-time classroom aide. Teachers were also randomly assigned to classes of different types. These assignments of students to class type were maintained through the third grade. Some students entered the study in the first grade and subsequent grades, but were randomly assigned to classes at that time. Teachers at each subsequent grade level were randomly assigned to classes as the experimental cohort passed through their grade. Districts had to agree to participate for four years, allow site visitations for verification of class sizes, interviewing, and data collection, including extra student testing. They also had to allow the research staff to randomly assign pupils and teachers to class types and to maintain the assignment of students to class types from Kindergarten through grade 3. The STAR project involves a rather broad range of schools from throughout a rather diverse state. It includes both large urban districts and small rural ones, and a range of wealth ranging from some of the wealthiest school districts in the state to some of the poorest. Thus, results obtained from the entire Project STAR sample are likely to be more generalizable than studies with more circumscribed samples.

How Persistent Are Teacher Effects?

Computing Teacher Effects

The analyses reported here make use of the Stanford Achievement Test (SAT) reading and mathematics test scores collected from Kindergarten through grade 3 as part of Project STAR. Because of the random assignment of students and teachers, the classes within each school are initially equivalent, and hence, any systematic differences in achievement among classes must be due to one of two sources: the treatment or differences in teacher effectiveness. Thus, within a school, any systematic differences in achievement between classrooms that had the same treatment must be due to variations in teacher effectiveness. In other words, in this randomized experiment the classroom mean residual adjusted for the treatment effect should be a good estimate of the teacher effect (e.g., teaching practice in a classroom) and it should be orthogonal to the treatment (see Raudenbush, 2004). In this case, it is reasonable to assume that this teacher effect is a causal effect. Because there are only a few classrooms in each school, we pool evidence of between-classroom within-school differences across schools. Following Nye and colleagues (2004) we operationalize the teacher effects as classroom-specific residuals that are adjusted for treatment effects (and student effects). Taking into consideration all these issues we compute the teacher effects as classroom-specific random effects (residuals) employing a three-level HLM (see Raudenbush & Bryk, 2002). 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 the teacher effects we used the same specification (see below) for mathematics and reading achievement for each grade (k to 3). Hence, for each grade the single-level equation for student i, in class j, in school k is

$$\begin{aligned} Y_{ijk} = \gamma_{000} + \gamma_{010} SMALLL_{jk} + \gamma_{020} AIDE_{jk} + \gamma_{100} FEMALE_{ijk} + \\ \gamma_{200} LOWSESi_{ijk} + \gamma_{300} MINORITY_{ijk} + \varepsilon_{ijk} + \xi_{0jk} + \eta_{00k} \end{aligned}$$

where Y_{ijk} represents student achievement in mathematics or reading in a specific grade, γ_{000} is the average achievement across students, classrooms, and schools, γ_{010} represents the overall small class effect, SMALL is a dummy variable for being in a small sized classroom, γ_{020} represents the overall regular class with a full-time aide effect, AIDE is a dummy variable for being in a regular sized classroom having a full-time classroom aide, γ_{100} is the overall gender effect, FEMALE is a dummy variable for gender, γ_{200} is the overall low SES effect, LOWSES is a dummy variable for free or reduced price lunch eligibility, γ_{300} is the overall minority effect, MINORITY is a dummy variable for minority group membership (indicating that the student was Black, Hispanic, or Asian), ε_{ijk} is a student-specific random effect (residual), ξ_{0jk} is a classroom-specific random effect (residual), and η_{00k} is a school-specific random effect (residual). In this model, the ξ_{0jk}'s represent the teacher effects adjusted for the effects of student gender, SES, minority group status, and treatment. Thus, the main objective of this analysis is to compute the between-classroom within-school residuals (the ξ_{0jk}'s). Note that for simplicity the estimates reported here are from a specification where only the classroom-specific and school-specific intercepts are treated as random at the second and third level respectively, but other analyses using additional random effects led to quite similar results. We conducted separate analyses for each of the two dependent variables, the SAT mathematics and reading test scores, for each of the three grade levels (1, 2, and 3). Therefore the analysis described here for achievement status was repeated six times.

We also conducted a residualized change analysis where we computed the teacher effects controlling for previous achievement as a covariate. In this case for each grade (1 and 2) the single-level equation for student i, in class j, in school k is

$$Y_{ijk} = \gamma^*_{000} + \gamma^*_{010}SMALL_{jk} + \gamma^*_{020}AIDE_{jk} + \gamma^*_{100}FEMALE_{ijk} + \gamma^*_{200}LOWSES_{ijk} + \gamma^*_{300}MINORITY_{ijk} + \gamma_{400}PRETEST_{ijk} + \varepsilon^*_{IJK} + \xi^*_{0JK} + \eta^*_{00K}$$

where γ_{400} represents the overall previous achievement effect, PRETEST is the achievement test in the previous year, and all other terms are as defined previously. Note that the asterisk indicates that the estimates are adjusted for the effects of previous achievement. The ξ^*_{0JK} ’s represent now the residualized teacher effects (adjusted for previous student achievement). We conducted separate analyses for each of the two dependent variables, the SAT mathematics and reading test scores, for each of the two grade levels (1 and 2). Therefore the analysis described here was repeated four times.

Assessing the Importance of Teacher Effects on Achievement Status

Once the teacher effects are computed the next series of analyses aims to gauge the impact of teacher effects on student achievement in subsequent years. The teacher effectiveness variable is now a student-specific attribute and hence a student predictor, and its estimate indicates whether the effectiveness of the teacher that a student received in one year affected that student’s achievement in the following years. The first part of this analysis examines whether the teacher effects influence student achievement in the following year. For example, we examined whether the teacher effects computed in Kindergarten (in mathematics and reading) are a significant (and important) predictor of student achievement in grade 1. To simplify interpretations, we standardized the outcomes (test scores) and the predictors included in the models to have a mean of zero and a standard deviation of unity in each grade. In particular, because of the standardization of all variables, all estimates are standardized regression coefficients. To determine the predictive efficacy of the teacher effects we also employed a three-level HLM. We used the same specification (see below) for mathematics and reading achievement for each grade (1 to 3). Hence, for each grade the single-level equation for student i, in class j, in school k is

$$Y_{ijk} = \gamma_{000} + \gamma_{100} FEMALE_{ijk} + \gamma_{200} LOWSES_{ijk} + \gamma_{300} MINORITY_{ijk} + \gamma_{400} TEACHEREFFECT_{ijk} + \varepsilon_{ijk} + \xi_{0jk} + \eta_{00k}$$

where Y_{ijk} represents student achievement in mathematics or reading in a specific grade (1, 2, 3), γ_{400} represents the overall teacher effect on student achievement in the following year, $TEACHEREFFECT$ is a continuous variable that represents the teacher effects in the previous year (k, 1, 2), and all other terms have been defined previously. The main objective of this analysis is to calculate the teacher effect estimate γ_{400}.

Notice that in this analysis it is important to compute the association between teacher effects in one year and student achievement in the following year controlling for the effects of the current teachers. That is, if the current teacher effects are correlated with teacher effects in previous years, and the teacher effects in the current year are not included in the equation, it is possible that the variable $TEACHEREFFECT$ is endogenous and that the estimate of γ_{400} is biased. The correlations among teacher effects in all four years (k to 3) ranged from 0.05 to 0.25. These correlations are rather low, which is expected because teachers were randomly assigned to classrooms within schools each year and the new students who participated in Project STAR in grades 1 to 3 for the first time were also randomly assigned to treatment conditions within schools. Nonetheless, the teacher effects are correlated and hence adjusting for teacher effects in the current year is important in order to obtain unbiased estimates of the teacher effects in the previous year on student achievement the following year. In a regression setting this can be achieved by introducing teacher fixed effects (as dummies) in the equation. In HLM this can be achieved by centering the level-1 (student) predictors at their classroom means (group mean centering). Notice that in the above equation all predictors are student-specific variables. The group-mean centering adjusts for teacher effects in the current year and it is equivalent to using teacher fixed effects in regression (Raudenbush & Wilms, 1995). Hence, in this analysis controlling for class size effects is not necessary (since it is a classroom variable).

The second part of this analysis examines the persistence of the teacher effects over time. Specifically, teacher effects in grades k and 1 were used simultaneously in the equation to predict student achievement in grade 2. In the same vein, teacher effects in grades k, 1, and 2 were used simultaneously in the equation to predict student achievement in grade 3. Below we portray the model for grade 3 achievement. The model for grade 2 achievement is similar, but the teacher effects in grade 2 are not included in the equation. Hence, for grade 3 for student i, in classroom j, in school k the three-level model is

$$Y_{ijk} = \gamma_{000} + \gamma_{100}FEMALE_{ijk} + \gamma_{200}LOWSES_{ijk} + \gamma_{300}MINORITY_{ijk} + \gamma_{400}TEACHEREFFECTK_{ijk} + \gamma_{500}TEACHEREFFECT1_{ijk} + \gamma_{600}TEACHEREFFECT2_{ijk} + \varepsilon_{ijk} + \xi_{0jk} + \eta_{00k}$$

where K, 1, and 2 indicate teacher effects in kindergarten, grade 1, and grade 2 respectively, $\gamma_{400}, \gamma_{500}, \gamma_{600}$ represent estimates of teacher effects in grades k, 1, and 2 respectively, and all other terms have been defined previously.

We also conduct analyses on achievement status using as predictors teacher effects that were adjusted for previous achievement as well (residualized teacher effects). That is, residualized (for achievement in grade k) teacher effects in grade 1 were used as predictors of student achievement in grade 2, and residualized (for achievement in grade 1) teacher effects in grade 2 were used as predictors of student achievement in grade 3. Finally, student achievement in grade 3 was regressed on residualized teacher effects in grades 1 and 2. Notice that this type of modeling resembles a value-added philosophy because teacher effects are adjusted for student achievement in the previous grade. In all analyses level-1 predictors were group-mean centered around their classroom means to adjust for current teacher effects.

Assessing the Importance of Teacher Effects on Achievement Gains

We also conduct analyses to examine teacher effects on achievement gains within the value-added framework. To that end, we determined how teacher effects in grades 1 and 2 affect grade 3 achievement controlling for student achievement in grade k (first year of the experiment). Hence, for grade 3 for student i, in classroom j, in school k the three-level model is

$$Y_{ijk} = \gamma^{*}_{000} + \gamma^{*}_{100}FEMALE_{ijk} + \gamma^{*}_{200}LOWSES_{ijk} + \gamma^{*}_{300}MINORITY_{ijk} + \gamma^{*}_{400}PRETESTK_{ijk} + \gamma^{*}_{500}TEACHEREFFECT1_{ijk} + \gamma^{*}_{600}TEACHEREFFECT2_{ijk} + \varepsilon^{*}_{jkl} + \xi^{*}_{0jk} + \eta^{*}_{00k}$$

where γ^{*}_{400} is the effect of achievement in kindergarten, *PRETESTK* represents achievement in kindergarten, and the asterisk indicates adjustment (due to previous achievement). Notice that in all models and specifications described here and above the teacher effects were modeled assuming a linear association with the outcomes.

Results

Teacher Effects on Achievement Status

The first part of the analysis examined the association between teacher effects in one year and student achievement in the following year for mathematics and reading achievement. The results of these analyses are reported in Tables 1 and 2. Specifically, Table 1 summarizes the associations between teacher effects in one year and student mathematics achievement in the following year controlling for gender, race, SES, and teacher effects in the current year. The lower panel of Table 1 portrays the association between residualized teacher effects in one year and student mathematics achievement the next year. The results for grade 1 indicate that a one standard deviation increase in teacher effectiveness in grade k corresponds to an increase of 0.070 standard deviations (SD) in mathematics achievement. This association is positive and significant and suggests that the teachers students receive in kindergarten affect their mathematics achievement in grade 1 (net of other effects in grade 1). The results for grades 2 and 3 are similar with positive and significant coefficients of larger magnitude (0.08 and 0.11 SD respectively). It is remarkable that the teacher effect in grade 3 is almost as large as the race effect in grade 3. The results for the residualized teacher effects also indicate positive and significant associations with student achievement but of somewhat smaller magnitude, which is expected because these teacher effects are adjusted for previous student achievement. The gender differences in mathematics are small and insignificant, while minority and low SES students have significantly lower mathematics achievement than their white and higher SES peers respectively.

Insert Table 1 Here

Table 2 summarizes the associations between teacher effects in one year and student reading achievement in the following year controlling for gender, race, SES, and teacher effects in the current year. Again, the lower panel of Table 2 portrays the association between residualized teacher effects in one year and student reading achievement the next year. The results for grade 1 indicate that a one standard deviation increase in teacher effectiveness in grade k corresponds to an increase of nearly 1/10 of a standard deviation in reading achievement. This association is positive and significant and suggests that the teachers students receive in kindergarten affect their reading achievement in grade 1 as well. The results for grades 2 and 3 are similar with positive and significant coefficients of larger magnitude (0.14 and 0.13 SD respectively). It is remarkable that the teacher effects in grades 1, 2, and 3 are larger than the race (and gender) effects in these grades. It is also interesting that the teacher effects are approximately one half as large as the low SES effects, which are typically substantial.

The results for the residualized teacher effects indicate positive and significant associations of somewhat smaller magnitude, which is expected because these teacher effects are also adjusted for previous student achievement. The gender differences in reading are positive and significant favoring female students, while minority and low SES students have significantly lower reading achievement than their white and higher SES peers respectively.

Insert Table 2 Here

Overall, these results show that the teacher effects in one year are positive and significant predictors of student achievement in the following year and are more pronounced in reading than in mathematics. On average these findings suggest that one standard deviation increase in teacher effects results in a 1/10 SD increase in student achievement, which is not trivial improvement. While this also holds true for the residualized teacher effects in reading, it does not in mathematics.

The second part of the analysis examined the persistence of the teacher effects over time. The objective of this analysis was to determine whether teacher effects are useful predictors of student achievement only in the following year or whether the teacher effects persist over time. The question here is do teacher effects in grades k and 1 persist to grade 3 when teacher effects in grades 2 and 3 are taken into account? Specifically, teacher effects in grades k and 1 were used simultaneously in the equation to predict student achievement in grade 2. Likewise, teacher effects in grades k, 1, and 2 were used simultaneously in the equation to predict student achievement in grade 3. Also, residualized teacher effects in grades 1, and 2 were used simultaneously in the equation to predict student achievement in grade 3.

The results of these analyses are reported in Table 3. Again all estimates are standardized regression coefficients. The results for grade 2 mathematics achievement indicate that teacher effects in grade k and 1 have independent, positive, and significant effects on mathematics student achievement. The results for grade 2 reading achievement are comparable, but much larger in magnitude. The results for grade 3 mathematics achievement indicate that teacher effects in grade k and in grade 2 have independent, positive, and significant effects on mathematics student achievement. However, teacher effects in grade 1 were not a significant predictor of grade 3 mathematics achievement in the presence of the other teacher effects in grades k and 2. The results for grade 3 reading achievement are similar, larger in magnitude, and suggest that the teacher effects in grade 1 are also a significant and positive predictor of reading achievement. It is noteworthy that the estimates of grade 1 and grade 2 teacher effects in reading are at least twice as large as those in mathematics. The results for residualized teacher effects on grade 3 mathematics and reading achievement were similar (positive, significant, and nearly as large).

Insert Table 3 Here

These results indicated that teacher effects persist over time up to grade 3 for both achievement tests. For example, grade k teacher effects were significant predictors of grade 2 and grade 3 achievement controlling for teacher effects in later grades. Grade 1 teacher effects were significant predictors of grade 2, but not of grade 3 achievement in mathematics. In reading, grade 1 teacher effects were consistently significant and larger than grade k teacher effects. Nonetheless, the teacher effects in the previous year (or in later years) were stronger than those in earlier years. Specifically, when modeling grade 3 achievement the estimates of the grade 2 teacher effects were larger than those in grades k and 1. Still, there is evidence that grade k and grade 1 teachers matter even in the presence of subsequent teacher effects, but these effects become smaller over time. Overall, these results indicate cumulative and independent teacher effects in grades k to 2. That is, the teacher effects in grade k persist to grade 3 in the presence of teacher effects in grades 1, 2, and adjusting for teacher effects in grade 3. In mathematics these results indicate that the achievement students who receive teachers in grades k to 2 who are quite effective (one standard deviation above the mean in the teacher effectiveness distribution) is increased by 1/5 SD, which is not trivial effect in educational research. The cumulative teacher effect for grade 3 reading achievement is even larger and nearly 1/3 SD, which is considerable.

Teacher Effects on Achievement Gains

The third part of the analysis examines the persistence of the teacher effects over time controlling for achievement in grade k. The objective of this analysis is to determine whether teacher effects persist over time when initial achievement is controlled for. This analysis addresses the question do teacher effects in grades 1 and 2 persist to grade 3 when achievement in grade k and teacher effects in grade 3 are taken into account? In particular, teacher effects in grades 1, and 2 were used simultaneously in the equation to predict student achievement in grade 3. The results of this analysis provide evidence of teacher effects on student learning between grade k and grade 3.

The results of this analysis are reported in Table 4. Again, all estimates are standardized regression coefficients. The results for grade 3 mathematics achievement indicate that teacher effects in grades 1 and 2 are positive but not significant at the .05 level. The first grade teacher effect estimate is close to zero, while the second grade teacher effect estimate is statistically insignificant but not trivial (nearly one half as large as the low SES effect). The results for grade 3 reading achievement show estimates of teacher effects that are positive, significant, and larger than those in mathematics. It is remarkable, that in reading the second grade teacher effect is larger than the low SES effect. Notice that these estimates of teacher effects are smaller than those reported in Table 3 for the residualized teacher effects. This suggests that grade k achievement has a considerable impact on teacher effects in subsequent years. Overall, these estimates point to important teacher effects on student learning in reading, while the evidence in mathematics is weaker. As before, the teacher effects persist in reading achievement, but the teacher effects in later years are larger than those in earlier years.

Insert Table 4 Here

Discussion

In this study we undertook the task to investigate whether teacher effects on student achievement persist over time using high-quality data from a four-year randomized experiment in which teachers and students were randomly assigned to classes. Our results suggest that teacher effects in early grades persist over time up to grade 3. Because of random assignment of teachers and students to classrooms in this experiment, our results provide stronger evidence about the durability of teacher effects. The findings support the idea that teachers *do matter* and significantly affect student achievement not only in the current or the following year, but in subsequent years as well. That is, we find that teacher effects are indeed cumulative. However, the results also show that teacher effects become smaller over time. This finding is consistent with those reported in previous studies that teacher effects diminish over time (e.g., McCaffrey, Lockwood, Mariano, & Satodji, 2005).

The findings on student achievement status suggest that the students who have very effective teachers at the 85th percentile of the teacher effectiveness distribution (assuming teacher effects are normally distributed) in three consecutive grades (k, 1, and 2) would experience achievement increases of about 1/5 SD in mathematics. The effects are similar in magnitude with the small class effects estimated using experimental data from Project STAR in any given year. The additive teacher effects are even more pronounced in reading. In particular, students who have very effective teachers at the 85th percentile of the teacher effectiveness distribution in three consecutive grades (k, 1, and

2) would experience achievement increases of about 1/3 SD in reading. These effects are considerable and comparable to achievement increases caused by small classes in early grades. Specifically, the cumulative effects of small classes (e.g., being in small classes for three or four years) are nearly 1/3 SD across achievement tests. Such effects in education are important and we argue that these cumulative teacher effects are certainly large enough to have policy significance. It is also noteworthy, that the teacher effects estimates in this study are typically larger than gender and race effects (which are not trivial, see e.g., Hedges & Nowell, 1995, 1999) and in certain cases nearly one half as large as the SES effects (which are typically substantial).

Although these effects are sizeable, we argue that they do not necessarily establish an upper bound on teacher effects. Specifically, our analyses may underestimate teacher effects because it is not clear that the outcome measures in Project STAR were strongly aligned with the intention of instruction. The effects of school inputs such as teacher effectiveness are expected to be the largest when the content covered during instruction is closely aligned with school outcomes such as student achievement measures (see, e.g., Walker and Schaffarzick, 1974; or Brimer et al., 1978). Recent work has further discussed the importance of alignment between tests and textbooks and classroom practices (see Porter, Smithson, Blank, & Zeidner, 2007).

It is interesting that the lasting benefits of teacher effects are consistently larger in reading than those in mathematics. Across all specifications the teacher effects estimates in reading were 25 to 50 percent larger than those in mathematics. This finding is somewhat inconsistent with that reported by Nye et al. (2004). Specifically, the authors found that for a specific (current) year the teacher effects were larger in mathematics than in reading. The results of the present study however, suggest that the durability of the teacher effects is more evident in reading than in mathematics at least in the early elementary school grades. This may be consistent with the notion that teachers put more emphasis on reading in early grades.

Overall, the findings reported here suggest that interventions to improve the effectiveness of teachers or identify effective teachers are promising strategies for improving student achievement for several years (at least in the early grades). The challenge then is to design and implement interventions such as professional development to improve teacher effectiveness. If there were a cost-effective intervention to improve teacher effectiveness, our findings suggest that the positive effects would be at least as large as those obtained from small class effects.

References

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.

Brimer, A., Madaus, F. G., Chapman, B., Kallaghan, T., & Wood, R. (1978). *Sources of difference in school achievement.* Windsor, Berks. SL4 1DF: NFER Publishing Company.

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.

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

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.

Hedges L. V., & Nowel, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*, 41-45.

Hedges, L. V., & Nowell, A. (1999). Changes in the Black-White gap in achievement test scores. *Sociology of Education, 72,* 111-135.

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

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.

McCaffrey, D. F., Lockwood, J. R., Mariano, L. T., & Sedodji, C. (2005). Challenges for value-added assessment of teacher effects. In R. Lissitz (Ed.), *Value-added models in education: Theory and applications* (pp 111-144). Maple Grove, MN: JAM Press.

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

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

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.

Porter, A. C., Smithson, J., Blank, R., Zeidner, T. (2007). Alignment as a teacher variable. *Applied Measurement in Education, 20,* 27-51.

Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, *20,* 307-335.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models*. Thousand Oaks, CA: Sage Publications.

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.

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.

Schmidt, W. H., Houang, R. T., & McKnight, C. C. (2005).Value-added research: Right idea but wrong solution? In R. Lissitz (Ed.), *Value-added models in education: Theory and applications* (pp 145-165). Maple Grove, MN: JAM Press.

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

Walker, D. F., & Schaffarzick, J. (1974). Comparing Curricula. *Review of Educational Research*, *44,* 83-111.

Table 1.
Standardized Coefficients of Teacher Effects in Mathematics: Grades 1 to 3

	Mathematics		
	Grade 1	Grade 2	Grade 3
Achievement Status			
Female	0.012	0.018	0.010
Minority	-0.191*	-0.139*	-0.109*
Low SES	-0.220*	-0.230*	-0.209*
Teacher Effect	0.070*	0.082*	0.105*
Female	-	0.011	0.017
Minority	-	-0.152*	-0.102*
Low SES	-	-0.216*	-0.223*
Residualized Teacher Effect	-	0.052*	0.077*

* $p < 0.05$

Note: Teacher Effect is Computed in the Previous Year.

Table 2.
Standardized Coefficients of Teacher Effects in Reading: Grades 1 to 3

	Reading		
	Grade 1	Grade 2	Grade 3
Achievement Status			
Female	0.106*	0.106*	0.099*
Minority	-0.060*	-0.089*	-0.097*
Low SES	-0.253*	-0.242*	-0.216*
Teacher Effect	0.095*	0.137*	0.126*
Female	-	0.101*	0.110*
Minority	-	-0.049	-0.103*
Low SES	-	-0.221*	-0.224*
Residualized Teacher Effect	-	0.126*	0.093*

* $p < 0.05$
Note: Teacher Effect is Computed in the Previous Year.

Table 3.
Standardized Coefficients of Cumulative Teacher Effects in Mathematics and Reading: Grades 2 and 3

	Mathematics	Reading
Grade 2		
Teacher Effect in Grade K	0.054*	0.083*
Teacher Effect in Grade 1	0.073*	0.130*
Grade 3		
Teacher Effect in Grade K	0.075*	0.082*
Teacher Effect in Grade 1	0.036	0.105*
Teacher Effect in Grade 2	0.084*	0.173*
Residualized Teacher Effect in Grade 1	0.037	0.116*
Residualized Teacher Effect in Grade 2	0.105*	0.149*

* $p < 0.05$

Table 4.
Standardized Coefficients of Teacher Effects in Mathematics and Reading in Grade 3: Achievement Gains Analysis

	Mathematics Grade 3	Reading Grade 3
Female	-0.011	0.063*
Minority	-0.011	-0.039
Low SES	-0.115*	-0.111*
Teacher Effect in Grade 1	0.008	0.063*
Teacher Effect in Grade 2	0.046	0.135*
Grade K Achievement	0.509*	0.479*

* $p < 0.05$, + $p < 0.1$