-- PRELIMINARY DRAFT, DO NOT QUOTE, COMMENTS WELCOME --

Value-Added Models and the Measurement of Teacher Quality

Douglas Harris
Dept. of Educational Leadership & Policy Studies
Florida State University
harris@coe.fsu.edu

Tim R. Sass
Dept. of Economics
Florida State University
tsass@coss.fsu.edu

Original Version: March 9, 2005

This Version: April 3, 2006

Abstract

The recent availability of administrative databases that track individual students and their teachers over time has lead to both a surge in research measuring teacher quality and interest in developing accountability systems for teachers. Existing studies employ a variety of empirical models, yet few studies explicitly state or test the assumptions underlying their models. Using an extensive database from the State of Florida, we test many of the central assumptions of existing models and determine the impact of alternative methods on measures of teacher quality. We find that the commonly used "restricted value-added" or "achievement-gain" model is a good approximation of the more cumbersome cumulative achievement model. Within the context of the restricted value-added model, we find it is important to control for unmeasured student, teacher and school heterogeneity. Relying on measurable characteristics of students, teachers and schools alone likely produces inconsistent estimates of the effects of teacher characteristics on student achievement. Moreover, individual-specific heterogeneity is more appropriately captured by fixed effects than by random effects; the random effects estimator yields inconsistent parameter estimates and estimates of time-invariant teacher quality that diverge significantly from the fixed effects estimator. In contrast, the exclusion of peer characteristics and class size each have relatively little effect on the estimates of teacher quality. Using aggregated grade-within-school measures of teacher characteristics produces somewhat less precise estimates of the impact of teacher professional development than do measures of the characteristics of specific teachers. Otherwise, aggregation to the grade level doesn't have a substantial effect. These findings suggest that many models currently employed to measure the impact of teachers on student achievement are mis-specified.

*We wish to thank the staff of the Florida Department of Education's K-20 Education Data Warehouse for their assistance in obtaining and interpreting the data used in this study. The views expressed is this paper are solely our own and do not necessarily reflect the opinions of the Florida Department of Education. This work is supported by Teacher Quality Research grant R305M040121 from the United States Department of Education Institute for Education Sciences. Thanks also go to Anthony Bryk for useful discussion of this research.

I. Introduction

In the last decade the availability of administrative databases that track individual student achievement over time has radically altered how education research is conducted and has brought fundamental changes to the ways in which educational programs and personnel are evaluated. Prior to the development of the Texas Schools Project by John Kain in the 1990s,[1] studies of student achievement and the role of teachers in student learning was limited largely to cross-sectional analysis of student achievement levels or simple two-period studies of student achievement gains. Now, in addition to Texas, statewide longitudinal databases exist in North Carolina and Florida as well as in large urban school districts such as New York, Chicago, Los Angeles and San Diego. The advent of these longitudinal databases has allowed researchers to measure changes in achievement at the individual student level, thereby controlling for the influences of students and families when evaluating educational programs.

The availability of student-level panel data is also fundamentally changing school accountability and the measurement of teacher performance. In Tennessee and Dallas, models of individual student achievement have been used to measure teacher performance.[2] While the stakes are currently low in these cases, there is growing interest among scholars and policymakers alike to use the measures for high-stakes merit pay, school grades, and other forms of accountability. Denver and Houston have recently adopted merit pay systems based on student performance and Florida plans to implement a statewide system beginning in the 2006-2007 school year.

The use of student-level longitudinal data in education research and systems of accountability is likely to expand even more rapidly in the coming years. With the new federal *No Child Left Behind* statute, testing requirements will increase so that all students will be tested in grades 3-8 in every state. Thus in a few years, all states will have the capability to track student achievement over time. This

[1] The Texas Schools Project was begun in 1992, but it took several years to create a unified database and the first research to exploit the data was not written until 1998.
[2] See Sanders and Horn (1998) and Mendro (1998) and references therein.

wealth of new data will bring great opportunities as well as significant challenges for the analysis of educational programs and policies.

In just the last few years, a plethora of studies have made use of the new student-level panel data sets to analyze the determinants of student achievement. However, no consensus has developed on the appropriate model specifications and empirical methods. In most cases the assumptions underlying the empirical models employed are unstated and untested and rarely are comparisons made between alternative methods.

Two recent studies, Todd and Wolpin (2005) and Ding and Lehrer (2005), investigate alternative forms of the cumulative achievement function, emphasizing the impact of historical home and schooling inputs on current achievement. Neither is directly concerned, however, with measuring the impact of teachers on student learning. Todd and Wolpin focus on the effect of family inputs on educational outcomes. Assignment of teachers to students within a school is assumed to be exogenous and only school-level averages of teacher inputs are used in their analysis. Ding and Lehrer exploit data from the Tennessee class-size experiment where students were randomly assigned to teachers and thus avoid the problems associated with measuring teacher quality.

In this paper we consider some of the same specification issues that are tested by Todd and Wolpin and Ding and Lehrer, but also investigate what factors are important in obtaining relatively consistent and precise estimates of the impact of teachers on student achievement. In section II we consider the general form of achievement functions and the effect of prior educational inputs on contemporaneous student achievement. Section III analyzes the measurement of schooling inputs that may influence student achievement, including peers, teachers and school-level variables. In section IV we discuss alternative methods of controlling for student and family characteristics. Section V discusses our data and in section VI we present our results. In the final section we summarize our findings and consider the implications for future research and for the implementation of accountability systems.

II. Achievement Model and the Treatment of Past Inputs

A. General Cumulative Model of Achievement

In order to clearly delineate the empirical models that have been estimated, we begin with a general cumulative model of student achievement in the spirit of Todd and Wolpin (2003):

$$\mathrm{A}_{\mathrm{it}} = \mathrm{A}_{\mathrm{t}}\left[\mathbf{X}_{\mathrm{i}}(\mathrm{t}), \mathbf{F}_{\mathrm{i}}(\mathrm{t}), \mathbf{E}_{\mathrm{i}}(\mathrm{t}), \mu_{\mathrm{i}0}, \varepsilon_{\mathrm{it}}\right] \tag{1}$$

where A_{it} is the achievement level for individual i at the end of their t^{th} year of life, $\mathbf{X}_{\mathrm{i}}(\mathrm{t})$, $\mathbf{F}_{\mathrm{i}}(\mathrm{t})$ and $\mathbf{E}_{\mathrm{i}}(\mathrm{t})$ represent the entire histories of individual, family and school-based educational inputs, respectively. The term $\mu_{\mathrm{i}0}$ is a composite variable representing time-invariant characteristics an individual is endowed with at birth (such as innate ability), and $\varepsilon_{\mathrm{it}}$ is a normally distributed, mean-zero error.

If we assume that the cumulative achievement function, $\mathrm{A}_{\mathrm{t}}[\cdot]$, does not vary with age[3] and is additively separable,[4] then we can rewrite the achievement level at age t as:

$$\begin{aligned}\mathrm{A}_{\mathrm{it}} = {} & \boldsymbol{\alpha}_1\mathbf{X}_{\mathrm{it}} + \boldsymbol{\alpha}_2\mathbf{X}_{\mathrm{it}-1} + \ldots + \boldsymbol{\alpha}_{\mathrm{t}}\mathbf{X}_{\mathrm{i}1} + \varphi_1\mathbf{F}_{\mathrm{it}} + \varphi_2\mathbf{F}_{\mathrm{it}-1} + \ldots + \varphi_{\mathrm{t}}\mathbf{F}_{\mathrm{i}1} + \boldsymbol{\beta}_1\mathbf{E}_{\mathrm{it}} + \boldsymbol{\beta}_2\mathbf{E}_{\mathrm{it}-1} + \ldots + \boldsymbol{\beta}_{\mathrm{t}}\mathbf{E}_{\mathrm{i}1} \\ & + \psi_{\mathrm{t}}\mu_{\mathrm{i}0} + \varepsilon_{\mathrm{it}}\end{aligned} \tag{2}$$

---

[3]This assumption implies that the impact of an input on achievement varies with the time span between the application of the input and measurement of achievement, but is invariant to the age at which the input was applied. Thus, for example, attending a private school in kindergarten has the same effect on achievement at the end of third grade as does attending a private school in second grade on fifth-grade achievement.

[4] Figlio (1999) explores the impact of relaxing the assumption of additive separability by estimating a translog education production function.

where α_1, φ_1 and β_1 represent the vectors of weights given to contemporaneous individual, family and school inputs, α_2, β_2 and φ_2 the weights given to last year's inputs and so on. The impact of the individual-specific time-invariant endowment in period t is given by ψ_t.

B. Cumulative Model with Fixed Family Inputs

Estimation of equation (2) requires data on both current and all prior individual, family and school inputs. However, administrative records contain only limited information on family characteristics and no direct measures of parental inputs.[5] Therefore, it is necessary to assume that family inputs are constant over time and are captured by a student-specific fixed component, ζ_i. However, the marginal effect of these fixed parental inputs on student achievement may vary over time and is represented by κ_t.

The assumption of fixed parental inputs of course implies that the level of inputs selected by families does not vary with the level of school-provided inputs a child receives. For example, it is assumed that parents do not systematically compensate for low-quality schooling inputs by providing tutors or other resources.[6] Similarly, it is assumed that parental inputs are invariant to achievement realizations; parents do not increase their inputs when their son or daughter does poorly in school.

The validity of the assumption that family inputs do not change over time is hard to gauge. Todd and Wolpin (2005), using data from the National Longitudinal Survey of Youth 1979 Child Sample (NLSY79-CS), consistently reject exogeneity of family input measures at a 90 percent confidence level, but not at a 95 percent confidence level. They have only limited

---

[5] Typically the only information on family characteristics is the student participation in free/reduced-price lunch programs, a crude measure of family income. Data in North Carolina also include teacher-reported levels of parental education.

[6] For evidence on the impact of school resources on parental inputs see Houtenville and Conway (2003) and BonesrØnning (2004).

aggregate measures of schooling inputs (average pupil-teacher ratio and average teacher salary measured at the county or state level) and the coefficients on these variables are typically statistically insignificant, whether or not parental inputs are treated as exogenous. Thus it is hard to know to what extent the assumed invariance of parental inputs may bias the estimated impacts of schooling inputs. It seems reasonable, however, that parents would attempt to compensate for poor school resources and therefore any bias in the estimated impacts of schooling inputs would be toward zero.

Given the assumption that family inputs and student ability are time-invariant (but may have differing marginal effects), we can combine the individual endowment and family inputs into a single component, $\omega_t\chi_i = \kappa_t\zeta_i + \psi_t\mu_{i0}$. The achievement equation becomes:

$$A_{it} = \boldsymbol{\alpha}_1\mathbf{X}_{it} + \boldsymbol{\alpha}_2\mathbf{X}_{it-1} + \ldots + \boldsymbol{\alpha}_t\mathbf{X}_{i1} + \boldsymbol{\beta}_1\mathbf{E}_{it} + \boldsymbol{\beta}_2\mathbf{E}_{it-1} + \ldots + \boldsymbol{\beta}_t\mathbf{E}_{i1} + \omega_t\chi_i + \varepsilon_{it} \quad (3)$$

Equation (3) represents our baseline model – the least restrictive specification of the cumulative achievement function that can conceivably be estimated with administrative data. In this very general specification current achievement depends on current and all prior individual time-varying characteristics and school-based inputs as well as the student's (assumed time invariant) family inputs and the fixed individual endowment.

Given the burdensome data requirements and computational cost of the full cumulative model, equation (3) has never been directly estimated for a large sample of students.[7] Rather, various assumptions have been employed to reduce the historical data requirements. There are several ways to avoid direct estimation of these lagged effects based on the assumed persistence

[7] Todd and Wolpin (2005) estimate the cumulative achievement model using a sample of approximately 7,000 students from the NLSY79-CS. Although they possess good measures of parental inputs and achievement levels the have only a few general measures of schooling inputs measured at the county or state level.

of lagged inputs on subsequent achievement. Below, we discuss the commonly used specifications and the associated restrictions on the cumulative achievement function, moving from the least-restrictive to the most restrictive specification.

C. The Unrestricted Value-Added Specification[8]

Suppose the marginal impacts of all prior school inputs decline geometrically with the time between the application of the input and the measurement of achievement at the same rate so that $\beta_2=\lambda\beta_1$, $\beta_3=\lambda^2\beta_1$, etc., where λ is a scalar. The achievement equation can then be expressed as:

$$A_{it} = \boldsymbol{\alpha}_1\mathbf{X}_{it} + \lambda\boldsymbol{\alpha}_1\mathbf{X}_{it-1} + \ldots + \lambda^{t-1}\boldsymbol{\alpha}_1\mathbf{X}_{i1} + \boldsymbol{\beta}_1\mathbf{E}_{it} + \lambda\boldsymbol{\beta}_1\mathbf{E}_{it-1} + \ldots + \lambda^{t-1}\boldsymbol{\beta}_1\mathbf{E}_{i1} + \omega_t\chi_i + \varepsilon_{it} \tag{4}$$

Taking the difference between current achievement and λ times prior achievement yields:

$$\begin{aligned} A_{it} - \lambda A_{it-1} = {} & \left[\boldsymbol{\alpha}_1\mathbf{X}_{it} + \lambda\boldsymbol{\alpha}_1\mathbf{X}_{it-1} + \ldots + \lambda^{t-1}\boldsymbol{\alpha}_1\mathbf{X}_{i1} + \boldsymbol{\beta}_1\mathbf{E}_{it} + \lambda\boldsymbol{\beta}_1\mathbf{E}_{it-1} + \ldots + \lambda^{t-1}\boldsymbol{\beta}_1\mathbf{E}_{i1} + \omega_t\chi_i + \varepsilon_{it}\right] - \\ & \left[\begin{array}{l} \lambda(\boldsymbol{\alpha}_1\mathbf{X}_{it-1}) + \lambda(\lambda\boldsymbol{\alpha}_1\mathbf{X}_{it-2}) + \ldots + \lambda(\lambda^{t-2}\boldsymbol{\alpha}_1\mathbf{X}_{i1}) + \lambda(\boldsymbol{\beta}_1\mathbf{E}_{it-1}) + \lambda(\lambda\boldsymbol{\beta}_1\mathbf{E}_{it-2}) + \ldots + \\ \lambda(\lambda^{t-2}\boldsymbol{\beta}_1\mathbf{E}_{i1}) + \lambda\omega_{t-1}\chi_i + \lambda\varepsilon_{it-1} \end{array}\right] \end{aligned} \tag{5}$$

Collecting terms, simplifying and adding λA_{it-1} to both sides produces:

$$A_{it} = \boldsymbol{\alpha}_1\mathbf{X}_{it} + \boldsymbol{\beta}_1\mathbf{E}_{it} + \lambda A_{it-1} + (\omega_t - \lambda\omega_{t-1})\chi_i + \varepsilon_{it} - \lambda\varepsilon_{it-1} \tag{6}$$

Assuming the impact of the initial endowment and family inputs on achievement, ω_t, changes at a constant rate then $(\omega_t - \lambda\omega_{t-1})$ can be expressed as a constant, ϖ. Combining the family inputs with the initial individual endowment into a single component yields:

[8] This specification is also sometimes referred to as the "covariate adjustment model" in the education literature.

$$A_{it} = \boldsymbol{\alpha}_1 \mathbf{X}_{it} + \boldsymbol{\beta}_1 \mathbf{E}_{it} + \lambda A_{it-1} + \gamma_i + \eta_{it} \tag{7}$$

where $\gamma_i = \varpi\chi_i$ is an individual student effect and $\eta_{it} = \varepsilon_{it} - \lambda\varepsilon_{it-1}$ is a random error.

Thus, given the assumed geometric rate of decay, the current achievement level is a function of contemporaneous student and school-based inputs as well as lagged achievement and an individual-specific effect. The lagged achievement variable serves as a sufficient statistic for all past schooling inputs, thereby avoiding the need for historical data on teachers, peers and other school-related inputs.

Ordinary least squares (OLS) estimation of equation (7) is problematic. Since η_{it} is a function of the lagged error, ε_{it-1}, the lagged achievement term, A_{it-1}, will be correlated with the error term in equation (7), η_{it}, and OLS estimates of equation (7) will in general be biased. A number of studies have estimated the effect of teacher quality on student achievement by estimating equation (7) by OLS, ignoring the correlation between the lagged dependent variable and error (eg. Aaronson, Barrow and Sander (2003), Clotfelter, Ladd and Vigdor (2005), Goldhaber and Brewer (1997) and Nye, Konstantopoulos and Hedges (2004)). To obtain consistent estimates it is necessary to use an instrumental variable estimation technique, typically incorporating A_{t-2} and longer lags as instruments. Because of the data requirements and computational burden this has rarely been done. Two exceptions are Ding and Lehrer (2005) and Sass (2006).

D. The Restricted Value-Added Specification

Rather than assume a constant rate of decay in the impact of schooling inputs on student achievement, an alternative approach is to assume that there is no decay in the effect of past schooling inputs on current achievement, ie. (1-λ)=0 or λ=1. Given this assumption the coefficient on lagged achievement in equation (7) is unity.[9] One can then subtract A_{it-1} from both sides of equation (7) to obtain:

$$\Delta A = A_{it} - A_{it-1} = \boldsymbol{\alpha}_1 \mathbf{X}_{it} + \boldsymbol{\beta}_1 \mathbf{E}_{it} + \gamma_i + \eta_{it} \tag{8}$$

As noted by Boardman and Murnane (1979) and Todd and Wolpin (2003), this implies that the effect of each input must be independent of when it is applied. In other words, school inputs each have an immediate one-time impact on achievement that does not decay over time. For example, the quality of a child's kindergarten must have the same impact on their achievement at the end of age 5 as it does on their achievement at age 18.

Equation (8), often called the "gain-score model" in the education literature, has been used by a number of authors to gauge the impact of teachers on student achievement.[10] As noted in Table 1, studies of teacher quality that employ achievement gains as the dependent variable include include Ballou, Sanders and Wright (2004), Goldhaber and Anthony (2004), Rivkin, Hanushek and Kain (2005) and Wright, Horn and Sanders (1997). None of these papers test the restriction on lambda, however.

E. The Contemporaneous Specification

A third alternative is to assume there is no effect of lagged inputs on current achievement. In other words, there is immediate and complete decay so that (1-λ)=1 or λ=0. In this case lagged achievement drops out of the achievement function and equation (7) becomes:

---

[9] Alternatively, the model can be derived by starting with a model of student learning gains (rather than levels) and assuming that there is no persistence of past schooling inputs on learning gains.

[10] This gain score model should be distinguished from the two-period gain score studies mentioned earlier that cannot control for the unobserved individual effects of students, teachers, and schools. See Harris and Sass (2006) for a discussion of the earlier type of gain score studies.

$$A_{it} = \boldsymbol{\alpha}_1 \mathbf{X}_{it} + \boldsymbol{\beta}_1 \mathbf{E}_{it} + \gamma_i + \eta_{it} \tag{9}$$

This is the approach taken by Dee (2004) and Rockoff (2004).

III. Modeling Teacher and School Inputs

A. Decomposing School Inputs

The vector of school-based educational inputs, $\mathbf{E}_{it}$, contains both school-level inputs such as the quality of principals and other administrative staff within school m, $\mathbf{S}_{mt}$, as well as a vector of classroom-level inputs in classroom j, $\mathbf{C}_{jt}$. The vector of classroom inputs can be divided into four components: peer characteristics, $\mathbf{P}_{-ijmt}$ (where the subscript –i students other than individual i in the classroom), time-varying teacher characteristics (such as experience), $\mathbf{T}_{kt}$ (where k indexes teachers), time-invariant teacher characteristics (such as "innate" ability and pre-service education), δ_k. and non-teacher classroom-level inputs (such as books, computers, etc.), $\mathbf{Z}_{jt}$. If we assume that, except for teacher quality, there is no variation in education inputs across classrooms within a school, the effect of $\mathbf{Z}_{jt}$ becomes part of the school-level input vector, $\mathbf{S}_{mt}$. If we further assume that school-level inputs are constant over the time span of analysis, they can be captured by a school fixed component, ϕ_m.[11] The value-added model can then be expressed as:

$$A_{it} = \boldsymbol{\beta}_1 \mathbf{X}_{it} + \boldsymbol{\beta}_2 \mathbf{P}_{-ijmt} + \boldsymbol{\beta}_3 \mathbf{T}_{kt} + \lambda A_{it-1} + \gamma_i + \delta_k + \phi_m + \eta_{it} \tag{10}$$

[11] Rarely do data exist on time-varying school-specific characteristics (ie. variables that change for one school but not for others within a district. One exception are the characteristics of school principals. We intend to analyze the impact of principals on student achievement in future work.

B. Modeling Teacher Effects

There are three major sources of variation in the modeling of teacher effects within educational production function models. First is the choice between measuring time-invariant teacher characteristics with the use of covariates, such as race, gender, and college selectivity or employing teacher-specific effects. Replacing the time-invariant teacher effect with a vector of constant teacher covariates $\mathbf{Y}_k$ in the value-added equation yields:

$$A_{it} = \boldsymbol{\beta}_1\mathbf{X}_{it} + \boldsymbol{\beta}_2\mathbf{P}_{-ijmt} + \boldsymbol{\beta}_3\mathbf{T}_{kt} + \lambda A_{it-1} + \gamma_i + \boldsymbol{\rho}\mathbf{Y}_k + \phi_m + \upsilon_{it} \tag{11}$$

where $\upsilon_{it} = (\delta_k - \boldsymbol{\rho}\mathbf{Y}_k) + \eta_{it}$. This approach will produce biased estimates if unmeasured time-invariant teacher characteristics, $(\delta_k - \boldsymbol{\rho}\mathbf{Y}_k)$, which are now part of the error term, are correlated with observed time-varying student, peer or teacher variables in the model (ie. $\mathbf{X}_{it}$, $\mathbf{P}_{-ijmt}$, $\mathbf{T}_{kt}$).

With teacher-specific effects, there is also the choice between fixed and random effects estimators. The impact of an individual teacher on student achievement can be modeled as a parameter specific to that teacher (a "fixed effect") or as a draw from a normal distribution. One can argue whether the population of teacher quality levels is normally distributed.[12] However, from a practical standpoint the choice between fixed and random effects often boils down to a tradeoff between computational cost and consistency of the estimator. Traditionally, the fixed effects or dummy-variable approach has been computationally burdensome. If student fixed effects are used to control for individual heterogeneity among students, then incorporating fixed effects for teachers has required incorporating a dummy variable for each teacher into the

[12] It is possible to estimate models which use other distributions for random effects, such as mixtures of normal distributions. See Verbeke and Lesaffre (1996).

model.[13] However, a new iterative fixed-effects estimator introduced by Arcidiacono et al. (2005) has greatly reduced the computational cost of estimating multi-level fixed effects models. This new technique has been employed by Harris and Sass (2006) to estimate models with student, teacher and school fixed effects using a large sample of students.

The random-effects approach essentially makes the teacher effect part of the error term and thus requires that the teacher effect be independent of the other explanatory variables in order to generate consistent estimates. Thus, for example, if peer effects are included in the model then random effects estimation would only yield consistent estimates if time-invariant teacher quality were uncorrelated with the characteristics of students in the classroom, which is unlikely. Since fixed effects are consistent (and random effects are inconsistent) when teacher effects are correlated with the explanatory variables in the model a Hausman test can be conducted to test for the consistency of the random effects estimator. To the best of our knowledge, the only existing study to conduct such as test is Goldhaber and Brewer (1997), who fail to reject the consistency of random teacher effects. In fact only five studies include any sort of teacher-specific effect along with a measure of student heterogeneity, three employing teacher fixed effects (Aaronson, Barrow and Sander (2003), Ballou, Sanders and Wright (2004) and Rivkin, Hanushek and Kain (2005))[14] and two utilizing teacher random effects (Goldhaber and Brewer (1997), Rockoff (2004)).[15]

The third specification issue related to teacher effects is whether the impact of teachers on student achievement varies across students. It may be that some teachers possess skills that

[13]Individual student effects can be taken into account by differencing the data with respect to student means, but then the teacher effects must be entered as dummy variables in a regression using the (student) de-meaned data. For a discussion of the computational issues involved see Andrews, Schank, and Upward (2004).

[14] Rivkin, Hanushek and Kain do not observe the specific classroom assignments of teachers and thus their teacher fixed effects are really school-by-grade effects that represent the average quality of teachers in a given grade level at a particular school.

[15] Teacher fixed effects have also been used in recent studies of peer effects. See Burke and Sass (2006) and Cooley (2005).

aid some students more than others or perhaps the racial/ethnic identity of a teacher has differential effects on students of different races and ethnicities. To control for potential variation in teacher effects among students a number of studies either include interactions between teacher and student characteristics (Wright, Horn and Sanders, (1997)) or conduct separate analyses for different groups of students (Aaronson, Barrow and Sander (2003), Dee (2004), Goldhaber and Anthony (2004)).

C. Modeling Peers, Other Classroom Factors, and School Effects

There is a rapidly growing empirical literature on classroom peer effects. It is well known that if students are assigned to classrooms non-randomly and peer-group composition affects achievement, then failure to control for the characteristics of classroom peers will produce biased estimates of the impact of teachers on student achievement. The measured teacher effects will capture not only the true impact of teachers but will also partially reflect the impacts of omitted peer characteristics. Recognizing this potential problem, the majority of the existing studies of teacher effects contain at least crude measures of peer characteristics like the proportion that are eligible for free/reduced-price lunch. An alternative approach is to focus on classes where students are, or appear to be, randomly assigned, as in Clotfelter, Ladd, and (2005), Dee (2004), and Nye, Konsantopoulos, and Hedges (2004).

As with the effects of peers, omission of other classroom-level variables can bias the estimated impact of teachers on student achievement if the allocation of non-teacher resources across classrooms is correlated with the assignment of teachers and students to classrooms. For example, principals may seek to aide inexperienced teachers by giving them additional computer resources. Similarly, classrooms containing a disproportionate share of low-achieving or disruptive students may receive additional resources like teacher aides. Due to the paucity of classroom data on non-teaching personnel and equipment, most studies omit any controls for non-teacher inputs. The only exceptions are Dee (2004) and Nye, Konstantopoulos and Hedges (2004) who use data from the Tennessee class-size experiment where classrooms were explicitly divided into three types: small classes, larger classes with an aide and larger classes with no aide.

Student achievement may be influenced by factors within schools but outside the classroom, such as the interactions and coordination among teachers, the leadership of the school principal, and the alignment of the adopted curriculum to achievement tests. It is rarely possible to measure any of these inputs directly and instead many researchers now include individual school effects. Like the teacher effects discussed above, these can be modeled as either fixed or random. When school-level effects are included, the teacher effects measure the influence of each teacher relative to the others within the same school.[16] This has important implications for the interpretation of our results, as discussed below.

D. Aggregation Issues

Historically, analyses of student achievement were often done at the school or even at the district level, since that was most disaggregate data available. With the advent of student-level panel data, individual student achievement can be analyzed, but frequently it is difficult to match students to their specific teacher. Precise student-teacher matching can be done at all grade levels in Florida and parts of New York City and for elementary-school students in North Carolina, it is not currently possible to link students to their teachers in other longitudinal databases from other areas, such as Texas.

[16]Teacher and school effects can be combined into a single teacher-school "spell" which then only requires a separate dummy for each unique teacher-school combination. Individual student effects can be taken into account by differencing the data with respect to student means, but then the teacher effects must be entered as dummy variables in a regression using the (student) de-meaned data. For a discussion of the computational issues involved see Andrews, Schank, and Upward (2004).

There are both potential advantages and disadvantages to aggregating data across teachers. As noted by Rivkin, Hanushek and Kain (2005), aggregating teacher characteristics to the grade within a school has the advantage of avoiding issues associated with non-random assignment of students to teachers. Also, it is well known that measurement error in the key independent variables places a downward bias on the coefficients (Grunfeld & Griliches, 1960). Measuring teacher attributes at the grade level rather than the individual level may therefore reduce this bias since errors at the individual teacher level may cancel out at the grade level. Hanushek, Rivkin and Taylor (1996) discuss this possibility, but they also show that aggregation can upwardly bias the estimated impacts of school resources in the presence of omitted variables, especially when the omitted variables occur at the same level as the aggregation. Aggregation will of course also tend to reduce the precision of estimates.

IV. Modeling Student/Family Inputs

The most important contributors to student learning are arguably the students themselves. Therefore, in the absence of random assignment, differences in average classroom performance will reflect not only the quality of teachers, but the ability of students as well. Consequently, it is important to control for student characteristics when evaluating the influence of teachers on student performance. While all studies include some measures of observed student characteristics, like race, gender and student mobility, there are distinct differences in how extant studies account for student and family characteristics that are typically unobservable, such as student ability and motivation.

A. Fixed and Random Student-Specific Effects

With the recent availability of longitudinal data, models of student achievement now frequently employ student-specific effects to control for time-invariant student and family characteristics. For example, in the economics literature Goldhaber and Anthony (2004), Rivkin, Hanushek and Kain (2005) and Rockoff (2004) directly capture the effect of time-invariant student-level ability and family inputs by incorporating individual fixed effects. In contrast to the use of student covariates, the fixed-effects approach should control for all time-invariant student characteristics, both observed and unobserved. In the education literature, few studies include individual-specific effects. One exception is Rowan, Correnti, and Miller (2002), who include student-specific effects in the context of a hierarchical linear model (HLM) framework.

As with the effects of teachers, the impact of unobserved student and family characteristics on student achievement can be modeled as either a fixed or random effect.[17] Random student effects will produce inconsistent estimates of model parameters if unobserved student heterogeneity is correlated with explanatory variables in the model. Since fixed effects yield consistent estimates even when they are correlated with other independent variables, the consistency of random effects estimators can be testing by comparing the parameter estimates from fixed and random-effects models via a Hausman test. Formal tests of the consistency of random effects estimators are rarely done, however,[18] perhaps due to the computational problems associated with estimating models with both student and teacher/school fixed effects models. As indicated above, however, recent advances in econometric methodology have greatly reduced the

[17] It is possible to estimate models which use other distributions for random effects, such as mixtures of normal distributions. See Verbeke and Lesaffre (1996).

[18] Todd and Wolpin (2005) conduct Hausman tests of fixed versus random student effects in their model of racial test score gaps and find that mother-specific endowment effects are not orthogonal to included inputs in the math achievement equation and thus the random effects specification is rejected. They do not reject the random effects specification for reading achievement, however.

computational cost of estimating multi-level fixed effects models, making this a viable alternative to the random effects methodology frequently employed in the past.

B. Measurable Student Inputs

Many studies in the literature on teacher quality estimate achievement models using observed time-invariant student and family characteristics, rather than student-specific effects to control for student ability and family inputs. Examples include Aaronson, Barrow and Sander (2003), Clotfelter, Ladd and Vigdor (2005) and Goldhaber and Brewer (1997). As with teacher covariates, the use of time invariant student characteristics like free lunch eligibility, race/ethnicity and disability status is potentially problematic. Any time-invariant student/family heterogeneity that is not captured by observed student characteristics becomes part of the error term. If the remaining unobserved student heterogeneity is correlated with observed time-varying student and school based independent variables in the model (ie. $\mathbf{X}_{it}$, $\mathbf{P}_{-ijmt}$, $\mathbf{T}_{kt}$), estimates of the model parameters will be biased. To minimize this problem, Clotfelter, Ladd and Vigdor (2005) analyze North Carolina classrooms with "apparent" random assignment of students (based on observed characteristics) to study the impact of teachers on student performance.

V. Data

In order to test alternative model specifications we utilize data come from the Florida Department of Education's K-20 Education Data Warehouse (EDW), an integrated longitudinal database covering all Florida public school students and school employees from pre-school through college. The EDW currently contains data for the 1995/1996 through 2003/2004 school years. Unlike most state-level administrative databases, the EDW includes not only test scores and demographic and programmatic information for individual students, but information on student enrollment, attendance and disciplinary actions as well. In addition, Florida's Education Data Warehouse incorporates employment records of all school personnel. Both the student and employee information can be linked to specific classrooms.

Although student records are available since the 1995/1996 school year, statewide standardized testing in consecutive grade levels did not begin in Florida until school-year 1999/2000. The state currently administers two sets of reading and math tests to all third through tenth graders in Florida. The "Sunshine State Standards" Florida Comprehensive Achievement Test (FCAT-SSS) is a criterion-based exam designed to test for the skills that students are expected to master at each grade level. The second test is the FCAT Norm-Referenced Test (FCAT-NRT), a version of the Stanford-9 achievement test. The Stanford-9 is a vertical or development-scale exam. Hence scores typically increase with the grade level and a one-point increase in the score at one place on the scale is equivalent to a one-point increase anywhere else on the scale. We use FCAT-NRT scale scores in all of the analysis. The use of the FCAT-NRT minimizes potential biases associated with "teaching to the test," since all school accountability standards, as well as promotion and graduation criteria in Florida are based on the FCAT-SSS, rather than the FCAT-NRT.

Although achievement test scores are available for both math and reading in grades 3-10, we limit our initial analysis to mathematics achievement in middle school, grades 6-8. We select middle-school mathematics classes for a number of reasons. First, it is easier to identify the relevant teacher and peer group for middle-school students than for elementary students. The overwhelming majority of middle school students in Florida move between specific classrooms for each subject whereas elementary school students typically receive the majority of their core academic instruction in a “self-contained” classroom. However, for elementary school students enrolled in self-contained classrooms, five percent are also enrolled in a separate math course and nearly 13 percent are enrolled in either special-education or gifted courses.

Second, parent “lobbying” and allocation of students to classrooms based on principals’ information regarding unmeasured student characteristics are more likely to lead to non-random classroom assignment in elementary school than in middle school. Since middle-school teachers often teach multiple sections of the same course, there is likely to be less parental pressure to have their child enrolled in a particular classroom (though they may still seek to have their child taught by a particular teacher). Also, because middle schools are larger, have more students per grade and students generally attend a different school for the preceding elementary grades, middle-school principals are less likely to possess information on unmeasured characteristics that can be used to make classroom assignments.

Third, because middle-school teachers often teach multiple sections of a course during an academic year, it is easier to clearly identify the effects of individual teachers on student achievement. In elementary school, teachers typically are with the same group of students all day long and thus teacher effects can only be identified by observing multiple cohorts of students taught by a given teacher over time. In contrast, both variation in class composition across sections at a point in time as well as variation across cohorts over time help to distinguish teacher effects from other classroom-level factors affecting student achievement in middle school.

We initially focus on math achievement rather than reading because it is easier to clearly identify the class and teacher most relevant to the material being tested. While some mathematics-related material might be presented in science courses, direct mathematics instruction almost always occurs in math classes. In contrast, middle school students in Florida may be simultaneously enrolled in "language arts" and reading courses, both of which may cover material relevant to reading achievement tests.

In addition to selecting middle-school math courses for analysis, we have limited our sample in other ways in an attempt to get the cleanest possible measures of classroom peers and teachers. First, we restrict our analysis of student achievement to students who are enrolled in only a single mathematics course (though all other students enrolled in the course are included in the measurement of peer-group characteristics). Second, to avoid atypical classroom settings and jointly taught classes we consider only courses in which 10-50 students are enrolled. Third, we eliminate any courses in which there is more than one "primary instructor" of record for the class. Finally, we eliminate charter schools from the analysis since they may have differing curricular emphases and student-peer and student-teacher interactions may differ in fundamental ways from traditional public schools.

Estimation of the achievement models with lagged test scores and individual fixed effects requires at least three consecutive years of student achievement data. Given statewide testing began in 1999/2000, our analysis is limited to Florida traditional public school students in grades 6-8 over the years 1999-2000 through 2003-2004 who took the FCAT-NRT for at least three consecutive years. This includes four cohorts of students, with over 120,000 students in each cohort. Unfortunately, it is not computationally tractable for us to consistently estimate models with lagged dependent variables using the entire sample. We therefore randomly select 100 middle schools for analysis, which results in approximately a twelve percent sample of the relevant population.[19]

[19] We randomly select 100 middle schools from all those operating in the 2002/2003 school year. However, we track students across all schools attended in the state and thus the number of schools in the sample exceeds 100.

VI. Results

A. The Value-Added Model and Persistence of Prior School Inputs

Recall from section II that lagged schooling inputs enter directly into the cumulative achievement function, but are captured by the lagged level of achievement in value-added formulation. Further, the restricted form of the value-added specification or "gain score" model assumes that the persistence of lagged schooling inputs is one. We present tests of these two assumptions in Table 2. In the first column of Table 2 we present estimates of the unrestricted value-added model (equation (10)) obtained by using the Arellano and Bond (1991) dynamic panel estimator. The coefficient on the lagged achievement score is not significantly different from zero, suggesting that past educational inputs do not affect current scores. However, this runs counter to previous work by Sass (2006), which estimates the coefficient on lagged achievement at 0.1 for mathematics and 0.2 for reading using data spanning grades 3-10.

We test the assumption that lagged achievement serves as a sufficient statistic for all past schooling inputs by adding twice-lagged measured inputs in the model (not shown in the table).[20] If lagged achievement does not capture the effects of prior inputs then past inputs should have statistically significant effects on achievement when added to the value-added model. Estimates of the unrestricted value-added model with twice-lagged schooling inputs are presented in the second column of Table 2. A Wald test on the joint significance of the lagged inputs fails to reject the null hypothesis, suggesting that lagged achievement serves as a sufficient statistic for historical schooling inputs.

[20] The twice lagged inputs included in the model were student mobility (number of schools attended, "structural" move, "non-structural" move), peer identity (proportion female, proportion black), class size and teacher experience (0 years, 1 year, 2-4 years).

To check the sensitivity of measures of teacher quality to assumptions about the persistence of lagged schooling puts we parametrically vary the persistence parameter, λ, from 0 to 1 in increments. A value of 1 corresponds to the gain-score model and 0 to the contemporaneous model. Estimates are presented in Tables 3A and 3B. The results in Table 3A indicate that the estimated impacts of time-varying teacher characteristics are similar across value-added specifications, but are qualitatively different for the contemporaneous specification. This is consistent with a recent study of charter schools by Sass (2006), which obtains qualitatively similar results for the restricted value-added (gain score) model and the unrestricted value-added model.

Table 3B presents correlations among the teacher effects that were estimated in the models displayed in Table 3A. Once again, the value-added models with different persistence levels all produce strikingly similar estimates. While the correlations decrease with the divergence in persistence assumptions, for values of λ from 0.2 to 1.0, the estimated teacher fixed effects are correlated at .88 and higher. Only for the contemporaneous model are the teacher effects much different than those from the gain-score model. In that case the correlation of teacher effects is only 0.76. This is consistent with the results of McCaffrey et al. (2004) who utilize two small-scale datasets to compare models: a simulation of 200 students and sample of 678 students from a single large suburban school district. McCaffey et al. (2004) find a high correlation between estimated teacher effects from models that include the restriction λ=1 versus one that leaves λ unrestricted.

Taking the results from Tables 3A and 3B together, it appears that use of the gain-score model to estimate teacher quality, rather than the unrestricted value-added model, should produce similar estimates. In the following analyses we utilize the gain score model and investigate how alternative specifications of the gain-score model impact estimates of teacher quality.

B. Alternative Measures of Time-Invariant Teacher Characteristics

Table 4 presents estimates of the impact of time-varying teacher characteristics on student achievement using alternative methods of controlling for time-invariant teacher attributes.[21] In the model presented in the first column of Table 4, teacher demographic characteristics (race, ethnicity, gender) are included as regressors while in the second column these variables are replaced with a set of teacher fixed effects.[22] Interestingly, with only teacher demographic characteristics both teacher experience and possession of an advanced degree are found to boost student achievement while these effects are statistically insignificant when unobserved teacher heterogeneity is taken into account with fixed effects. These results suggest that the use of teacher fixed effects is important to adequately control for unmeasured aspects of teacher quality.[23]

B. Differing Controls for Classroom and School Characteristics

Table 5A presents estimates of the impact of time-varying teacher characteristics on student achievement from models with differing controls for peer influences, class size and school characteristics. Inclusion/exclusion of variables to control for peer quality, class size and school quality has virtually no effect on the estimated impacts of teacher experience, professional development and advanced degrees. At first blush this might seem surprising. However, it is important to recognize that all of the estimated models include teacher fixed effects and thus

---

[21] Unlike the previous set of tables, there are no correlations of teacher effects since only one model include teacher fixed effects.
[22] Ideally, one would want to include measures of pre-service ability and training, such as college entrance exam scores, college coursework, etc. in the vector of time-invariant teacher characteristics. Unfortunately, we currently possess this information for only a small fraction of Florida teachers.
[23] In contrast, Hanushek (1992) obtains similar results when comparing teacher covariates with teacher fixed effects.

control for unobserved teacher quality that might be correlated with student and school characteristics. If the models had excluded teacher fixed effects we would expect significant variation in the estimated coefficients across the various models.

In Table 5B, the correlations in the estimated teacher effects from alternative models clearly indicate that inclusion/exclusion of school effects greatly impacts estimated teacher effects while controls for class size and peer characteristics have only minor impacts. This is consistent with the notion that there is significant sorting in teacher quality across schools. If unobserved teacher quality is correlated with school quality then removing school fixed effects would greatly alter the estimated teacher effects, which is what we observe. Put differently, if teachers are not randomly assigned across schools, then the performance of a teacher relative to her average colleague within a school (ie. the fixed teacher effect when school effects are included) will be different than her performance relative to the average teacher in the school system (ie. the fixed teacher effect when school effects are excluded).

Our findings regarding the importance of school effects are consistent with the results of Aaronson, Barrow and Sander (2003). They find that the exclusion of school fixed effects can be easily rejected by an *F*-test. Similarly, McCaffrey, et al. (2004) find that when school effects are excluded, estimated teacher effects are negatively correlated with the proportion of students receiving free and reduced-price lunch. When school fixed effects are included this correlation is essentially eliminated. Further, with school fixed effects the within-school variance in teacher effects is much smaller than in a model with school-specific controls.

C. Differing Controls for Student Characteristics

Table 6A presents estimates of the restricted value-added model with three alternative measures of student heterogeneity: time-invariant student characteristics[24], student fixed effects and student random effects. Similar to the results for time-invariant teacher characteristics, we find the use of covariates rather than fixed effects to capture student heterogeneity greatly alters the estimated impacts of time-varying teacher characteristics. Just as when teacher covariates are employed, using student covariates rather than fixed effects suggests that teacher experience significantly influences student outcomes whereas the impact of experience vanishes when student fixed effects are included in the model. This suggests that unmeasured student ability is correlated with both measured teacher characteristics (eg. experience, professional development) and unmeasured teacher attributes (eg. pre-service ability and training), as one would expect if students are not randomly assigned to teachers. Similarly, Aaronson, Barrow, and Sander (2003) easily reject student covariates as a substitute for individual (fixed) student effects.

The third column of Table 6A presents estimates of the gain-score model when student random effects are used to capture student heterogeneity. The results are strikingly similar to those from the model with student covariates and quite different from those from the student-fixed effects model. The fourth column presents the difference between the coefficients from the fixed and random effects models. A Hausman test on the model parameters (other than the student/teacher/school effects) yields a chi-squared value of 189.76 with 29 degrees of freedom. Thus we can clearly reject the null hypothesis that random effects are uncorrelated with the explanatory variables in the model at better than a 99 percent confidence level. Thus our findings indicate that the random effects estimator produces inconsistent parameter estimates in

[24] The measured student characteristics are race/ethnicity, foreign/native born, language parents speak at home and free lunch status.

the context of a gain-score model. This is a significant finding given the extensive use of hierarchical linear modeling (HLM) which use random effects for both students and teachers. Our results suggest that these models may suffer from considerable bias.

The differences between the fixed and random effects estimators are mirrored in the correlation matrices presented in Table 6B. The estimated teacher fixed effects vary substantially with how student heterogeneity is modeled. The correlation between the estimated teacher effects from the two models is only 0.39. Given that the student fixed effects estimator is always consistent, the low correlation provides further evidence that the random effects specification produces inconsistent parameter estimates.

D. Data Aggregation

Table 7 presents evidence on the effects of data aggregation. Both columns in Table 7 provide estimates of gain-score models that include student fixed effects. The results in the first column are from a model which measures the characteristics of individual teachers and which includes teacher and school fixed effects. The second column presents estimates from a model that uses grade-by-school-by-year average teacher characteristics instead of the characteristics of specific teachers and employs school-by-year and grade-by-school fixed effects (rather than student and school fixed effects), similar to Rivkin, Hanushek and Kain (2005). The estimates from the aggregate and disaggregate models are similar in many respects, though the two models differ in the estimated impacts of teacher professional development. The aggregate model finds contemporaneous total professional development hours to have a positive impact on student achievement and lagged total professional development to have no effect. In contrast, estimates from the disaggregated model indicates that contemporaneous total professional development hours are statistically insignificant and lagged hours have are negatively correlated with student achievement gains. Likewise, twice-lagged content-based professional development for teachers is found to boost test scores in the disaggregated model but not in the aggregate model.

V. Conclusion

Past research on value-added modeling has been significantly hampered by data limitations, which, in turn, has forced researchers to estimate mis-specified models. The data we use from Florida avoid these limitations and allow for thorough testing of model assumptions and their impact on estimates.

Our results suggest that student and teacher heterogeneity are the most important issues that value-added models must contend with. We confirm the finding of past studies that covariates are inadequate replacements for individual student and teacher effects. Moreover, random effects models yield inconsistent estimates of model parameters due to correlation between the random effects and explanatory variables in the model. The biases introduced by covariate and random effects models extend both to the estimates of the unobserved teacher quality and the effects of time-varying teacher characteristics (experience and professional development) on student achievement.

We also reject the exclusion of individual school effects. There is a low correlation between individual teacher effects from models with and without school effects, suggesting that estimated teacher effects partly reflect the influence of school-wide inputs when school effects are omitted.

The modeling of students' peers and other non-teacher classroom-level factors appear to have relatively little impact on the estimated effects of teacher quality. The same is true of the modeling of lagged school inputs. We also find that the assumed persistence of educational inputs makes little difference, suggesting the choice between simple gain-score models and unrestricted value-added models, may not be very important. We also find that prior test scores serve as a sufficient statistic for past educational inputs, indicating one can utilize value-added models rather than the more cumbersome cumulative models of achievement.

These results have significant implications for both educational research and policy. First, the importance of individual fixed effects calls into question the common assumptions made by educational researchers who use HLM analysis. This includes the current accountability systems in Dallas and Tennessee which are also based on an HLM framework. But perhaps the most significant problems in using value-added models for accountability are that school effects appear to play an important role and that teachers are non-randomly assigned to schools. The first finding implies that, if school effects are excluded from the models, then the teacher effects are biased and capture factors that appear to be outside the control of the teachers. However, the second fact means that, if the school effects are included in the models, then it is possible only to compare teachers within schools, which may create unproductive competition between teachers. Thus, there appears to be a fundamental trade-off between these two approaches for the purposes of accountability.

The implications of our results for research on teacher quality are somewhat clearer. By testing the assumptions of past models, we have narrowed the range of justifiable models as well as the data requirements that must be met in order to estimate them. Given the coming expansion of standardized testing and improved database capabilities, the importance of understanding value-added modeling will only continue to grow.

Bibliography

Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster and Josh Kinsler (2005). "Estimating Spillovers in the Classroom with Panel Data," unpublished manuscript.

Andrews, Martyn, Thorsten Schank and Richard Upward, "Practical Estimation Methods for Linked Employer-Employee Data," unpublished manuscript (2004).

Aaronson, Daniel, Lisa Barrow, and William Sander, "Teachers and Student Achievement in the Chicago Public High Schools," unpublished manuscript (2003).

Arellano, Manuel, and Stephen Bond. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58 (1991), 277-297.

Ballou, Dale, "Rejoinder," *Journal of Educational and Behavioral Statistics*, 29:1 (2004), 131-134.

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

Boardman, Anthony E., and Richard J. Murnane, "Using Panel Data to Improve Estimates of the Determinants of Educational Achievement," *Sociology of Education*, 52 (1979), 113-121.

Bonesrønning, Hans, "The Determinants of Parental Effort in Education Production: Do Parents Respond to Changes in Class Size?," *Economics of Education Review*, 23 (2004), 1-9.

Burke, Mary A., and Tim R. Sass, "Classroom Peer Effects and Student Achievement," unpublished manuscript (2004).

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor, "Teacher-Student Matching and the Assessment of Teacher Effectiveness," unpublished manuscript (2005).

Cooley, Jane, "Desegregation and the Achievement Gap: Do Diverse Peers Help?," unpublished manuscript (2005).

Dee, Thomas S., "Teachers, Race and Student Achievement in a Randomized Experiment," *Review of Economics and Statistics*, 86:1 (2004), 195-210.

Ding, Weili, and Steven F. Lehrer, "Accounting for Unobserved Ability Heterogeneity within Education Production Functions," unpublished manuscript (2005).

Figlio, David N., "Functional Form and the Estimated Effects of School Resources," *Economics of Education Review*, 18 (1999), 241-252.

Goldhaber, Dan, and Emily Anthony, "Can Teacher Quality be Effectively Assessed?," unpublished manuscript (2004).

Goldhaber, Dan D., and Dominic J. Brewer, "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity," *Journal of Human Resources*, 32:3 (1997), 505-523.

Grunfeld, Yehuda and Zvi Griliches, "Is Aggregation Necessarily a Bad Thing," *Review of Economics and Statistics*, 42 (1960), 1-13.

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

Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor, "Aggregation and the Estimated Effects of School Resources," *Review of Economics and Statistics*, 78:4 (1996), 611-627.

Harris, Douglas and Tim R. Sass, "The Effects of Teacher Training on Teacher Value Added," unpublished manuscript (2006).

Houtenville, Andrew J. and Karen S. Conway, "Parental Effort, School Resources and Student Achievement: Why Money May Not 'Matter'," unpublished manuscript (2003).

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

Mendro, Robert L., "Student Achievement and School and Teacher Accountability," *Journal of Personnel Evaluation in Education*, 12:3 (1998), 257-267.

Nye, Barbara, Spyros Konstantopoulos, and Larry V. Hedges, "How Large are Teacher Effects?," *Educational Evaluation and Policy Analysis*, 26:3 (2004), 237-257.

Raundenbush, Stephen W., "What are Value-Added Models Estimating and What Does This Imply for Statistical Practice?," *Journal of Educational and Behavioral Statistics*, 29:1 (2004), 121-129.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain, "Teachers, Schools and Academic Achievement," *Econometrica*, 73:2 (2005), 417-458.

Rockoff, Jonah E., "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *American Economic Review*, 94:2 (2004), 247-252.

Rowan, Brian, Richard Correnti, and Robert J. Miller, "What Large-Scale, Survey Results tell us About Teacher Effects on Student Achievement: Insights from the Prospects Study of Elementary schools," *Teachers College Record*, 104:8 (2002), 1525-1567.

Sanders, William L., and Sandra P. Horn, "Research Findings From the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research," *Journal of Personnel Evaluation in Education*, 12:3 (1998), 247-256.

Sass, Tim R., "Charter Schools and Student Achievement in Florida," *Education Finance and Policy*, 1:1 (2006), 91-122.

Todd, Petra E. and Kenneth I. Wolpin, "On the Specification and Estimation of the Production Function for Cognitive Achievement," *The Economic Journal*, 113 (2003), F3-F33.

Todd, Petra E., and Kenneth I. Wolpin, "The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps," unpublished manuscript (2005).

Verbeke, Geert, and Emmanuel Lesaffre, "A Linear Mixed-Effects Model with Heterogeneity in the Random-Effects Population," *Journal of the American Statistical Association*, 91:433 (1996), 217-221.

Wright, S. Paul, Sandra P. Horn, and William L. Sanders, "Teacher and Classroom Context Effects on Student Achievement: Implications for Teacher Evaluation," *Journal of Personnel Evaluation in Education*, 11 (1997), 57-67.

Table 1
Summary of Studies and Models

Study	Teacher Controls	Teacher Effects Vary by Student Type?	Student Controls	School Controls	Persistence of Schooling Inputs	Controls for Non-Teacher Classroom Inputs?	Controls for Peer effects?	Class Size Control?
Aaronson, Barrow and Sander (2003)	Fixed Effects, Covariates	Yes	Covariates	Fixed Effects	Constant rate of decay	No	Absences, Lagged Test Score	Yes
Ballou, Sanders and Wright (2004)	Fixed Effects	No	Covariates	No	No decay	No	Grade-Level Free Lunch	No
Clotfelter, Ladd and Vigdor (2004)	"Apparent" Random Assignment, Covariates	No	"Apparent" Random Assignment, Covariates	Fixed Effects	Constant rate of decay	No	No	Yes
Dee (2004)	Random Assignment, Covariates	Yes	Random Assignment, Covariates	Fixed Effects	Complete decay	Yes	Free Lunch, Race, Gender	Yes
Goldhaber & Brewer (1997)	Covariates, Fixed Effects, Random Effects	No	Covariates	Covariates, Fixed Effects	Constant rate of decay	No	Race	Yes
Goldhaber and Anthony (2004)	Covariates	Yes	Fixed Effects, Covariates	Fixed Effects, Covariates	No decay	No	School-level Free-Lunch, Race	Yes (school-level)
Hanushek (1992)	Fixed Effects, Covariates	No	Covariates	No	No decay	No	No	Yes
Nye, Konstantopoulos, & Hedges (2004)	Random Assignment, Covariates	No	Random Assignment, Covariates	Random Effects	Constant rate of decay	Yes	No	Yes
Rivkin, Hanushek and Kain (2005)	Fixed Effects (by grade level), experience	No	Fixed Effects	Fixed Effects	No decay	No	No	No
Rockoff (2004)	Random Effects, Experience	No	Fixed Effects	Fixed Effects	Complete decay	No	Lagged Test Score	Yes
Rowan, Correnti and Miller (2002)	Covariates	Yes	Time-Varying Random Effects	Random Effects	Constant Rate of Decay, No decay	No	No	
Wright, Horn and Sanders (1997)	Random Effects	Yes	None	No	No decay	No	Test Scores	Yes

Table 2
Estimates of the Impact of Teacher Characteristics on Student Math Achievement Using Unrestricted Value-Added Model (100 Florida Middle Schools, 1999/00-2003/04)

Explanatory Variable	Without Lagged Inputs	With Lagged Inputs
Achievement Score$_{t-1}$	-0.0337	-0.0371
	(1.21)	(1.34)
0 Years of Experience	-2.3451	-1.7784
	(1.51)	(1.13)
1 Year of Experience	-0.0952	0.1887
	(0.10)	(0.19)
2-4 Years of Experience	-2.8559***	-2.9996***
	(3.05)	(3.21)
Total In-service Hours$_t$	0.0166**	0.0170**
	(2.27)	(2.32)
Total In-service Hours$_{t-1}$	0.0024	0.0034
	(0.28)	(0.39)
Total In-service Hours$_{t-2}$	0.0001	0.0008
	(0.02)	(0.01)
Total In-service Hours$_{t-3}$	-0.0057	-0.0058
	(0.93)	(0.95)
Content In-service Hours$_t$	-0.0381***	-0.0385***
	(3.21)	(3.26)
Content In-service Hours$_{t-1}$	-0.0208	-0.0204
	(1.51)	(1.48)
Content In-service Hours$_{t-2}$	0.0081	0.049
	(0.67)	(0.40)
Content In-service Hours$_{t-3}$	0.0132	0.0140
	(0.91)	(0.96)
Advanced Degree	-1.1845	-1.3680*
	(1.58)	(1.83)
Student Fixed Effects	Yes	Yes
Teacher Fixed Effects	Yes	Yes
School Fixed Effects	Yes	Yes
Chi-Squared Test on Lagged Inputs		7.51
Number of Students (after first differencing)	7,816	7,816
Number of Observations (after first differencing)	10,329	10,329

Models are estimated using the Arellano and Bond dynamic panel data estimator. Models include the following time varying student/class characteristics: number of schools attended by the student in the current year, “structural” move by student, “non-structural” move by student, indicator of a student repeating a grade, class size, fraction of classroom peers who are female, fraction of classroom peers who are black, average age (in months) of classroom peers, fraction of classroom peers who changed schools, fraction of classroom peers who made a “structural move.” All models also include year, grade level, and repeater-by-grade dummies. Absolute values of t-statistics appear in parentheses. * indicates statistical significance at the .10 level and ** indicates significance at the .05 level and *** indicates significance at the .01 level in a two-tailed test.

Table 3A
Estimates of the Impact of Teacher Characteristics on Student Math Achievement Using Restricted Value-Added Models With Differing Assumptions Regarding Persistence of Schooling Inputs (100 Florida Middle Schools, 1999/00-2003/04)

Explanatory Variable	Restricted Value Added λ=1	Value-Added (Partial Decay) λ=.8	λ=.6	λ=.4	λ=.2	Contem-poraneous λ=0
0 Years of Experience	-2.7495	-2.2506	-1.7517	-1.2529	-0.7540	-0.2551
	(0.91)	(0.83)	(0.72)	(0.57)	(0.38)	(0.14)
1 Year of Experience	1.7714	1.8542	1.9369	2.0197	2.1025	2.1852*
	(0.82)	(0.96)	(1.11)	(1.29)	(1.48)	(1.66)
2-4 Years of Experience	-0.4575	-0.1030	0.2515	0.6061	0.9606	1.3152
	(0.29)	(0.07)	(0.20)	(0.52)	(0.91)	(1.36)
Total In-service Hours$_t$	0.0115	0.0113	0.0110	0.0107	0.0105	0.0102
	(1.07)	(1.16)	(1.27)	(1.38)	(1.49)	(1.59)
Total In-service Hours$_{t-1}$	-0.0240**	-0.0216**	-0.0191**	-0.0167**	-0.0143**	-0.0118*
	(2.22)	(2.21)	(2.19)	(2.12)	(2.00)	(1.79)
Total In-service Hours$_{t-2}$	-0.0225**	-0.0200**	-0.0176**	-0.0152*	-0.0127*	-0.0103
	(2.06)	(2.04)	(2.00)	(1.93)	(1.79)	(1.57)
Total In-service Hours$_{t-3}$	-0.0155	-0.0143	-0.0131	-0.0118	-0.0106	-0.0094
	(1.53)	(1.56)	(1.59)	(1.60)	(1.58)	(1.49)
Content In-service Hours$_t$	-0.0177	-0.0187	-0.1097	-0.0207	-0.0217*	-0.0227**
	(0.92)	(1.08)	(1.28)	(1.50)	(1.74)	(1.99)
Content In-service Hours$_{t-1}$	0.0216	0.0183	0.0150	0.0117	0.0084	0.0051
	(1.07)	(1.01)	(0.92)	(0.80)	(0.63)	(0.41)
Content In-service Hours$_{t-2}$	0.0517**	0.0434**	0.0352**	0.0269*	0.0186	0.0104
	(2.37)	(2.21)	(2.00)	(1.70)	(1.30)	(0.78)
Content In-service Hours$_{t-3}$	0.0100	0.0084	0.0068	0.0052	0.0035	0.0019
	(0.42)	(0.39)	(0.35)	(0.30)	(0.23)	(0.13)
Advanced Degree	0.4133	0.5187	0.6241	0.7295	0.8350	0.9404
	(0.24)	(0.34)	(0.45)	(0.59)	(0.75)	(0.93)
Student Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of Students	47,442	47,442	47,442	47,442	47,442	47,442
Number of Observations	74,196	74,196	74,196	74,196	74,196	74,196

Models include the following time varying student/class characteristics: number of schools attended by the student in the current year, "structural" move by student, "non-structural" move by student, indicator of a student repeating a grade, class size, fraction of classroom peers who are female, fraction of classroom peers who are black, average

age (in months) of classroom peers, fraction of classroom peers who changed schools, fraction of classroom peers who made a "structural move." All models also include year, grade level, and repeater-by-grade dummies. Absolute values of t-statistics appear in parentheses. * indicates statistical significance at the .10 level and ** indicates significance at the .05 level and *** indicates significance at the .01 level in a two-tailed test.

Table 3B
Correlation of Estimated Teacher Fixed Effects
From Student Math Achievement Restricted Value-Added Models
With Differing Assumptions Regarding Persistence of Schooling Inputs
(100 Florida Middle Schools, 1999/00-2003/04)

Restricted Value Added	Value-Added (Partial Decay)				Contem-poraneous
λ=1	λ=.8	λ=.6	λ=.4	λ=.2	λ=0
1.0000					
0.9967	1.0000				
0.9828	0.9946	1.0000			
0.9492	0.9717	0.9910	1.0000		
0.8813	0.9169	0.9534	0.9852	1.0000	
0.7612	0.8115	0.8679	0.9266	0.9773	1.0000

Table 4
Estimates of the Impact of Teacher Characteristics on Student Math Achievement Using Restricted Value-Added Models With Differing Controls for Teacher Heterogeneity (100 Florida Middle Schools, 1999/00-2003/04)

Explanatory Variable	Time-Invariant Teacher Characteristics	Teacher Fixed Effects
0 Years of Experience	-3.8087***	-1.5874
	(3.62)	(0.49)
1 Year of Experience	-0.4017	2.4501
	(0.58)	(1.07)
2-4 Years of Experience	-2.3244***	0.1362
	(3.56)	(0.08)
Total In-service Hours$_t$	0.0074	0.0096
	(1.25)	(0.85)
Total In-service Hours$_{t-1}$	-0.0172***	-0.0257**
	(2.81)	(2.27)
Total In-service Hours$_{t-2}$	0.0005	-0.0208*
	(0.08)	(1.82)
Total In-service Hours$_{t-3}$	0.0048	-0.0135
	(0.80)	(1.25)
Content In-service Hours$_t$	-0.0104	-0.0153
	(0.91)	(0.76)
Content In-service Hours$_{t-1}$	0.0151	0.0210
	(1.32)	(1.00)
Content In-service Hours$_{t-2}$	0.0116	0.0511**
	(0.93)	(2.16)
Content In-service Hours$_{t-3}$	0.0018	0.0127
	(0.11)	(0.47)
Advanced Degree	1.2414**	0.0812
	(2.45)	(0.04)
Student Fixed Effects	Yes	Yes
Teacher Fixed Effects	No	Yes
School Fixed Effects	Yes	Yes
Number of Students	45,914	45,914
Number of Observations	70,437	70,437

Models include the following time varying student/class characteristics: number of schools attended by the student in the current year, "structural" move by student, "non-structural" move by student, indicator of a student repeating a grade, class size, fraction of classroom peers who are female, fraction of classroom peers who are black, average age (in months) of classroom peers, fraction of classroom peers who changed schools, fraction of classroom peers who made a "structural move." All models also include year, grade level, and repeater-by-grade dummies.

Absolute values of t-statistics appear in parentheses. * indicates statistical significance at the .10 level and ** indicates significance at the .05 level and *** indicates significance at the .01 level in a two-tailed test.

Table 5A
Estimates of the Impact of Teacher Characteristics on Student Math Achievement Using Restricted Value-Added Models With Differing Controls for Classroom and School Characteristics (100 Florida Middle Schools, 1999/00-2003/04)

Explanatory Variable	Peer Var., Class Size, School F.E. Included	No. Peer Var.	No Class Size	No School F.E.	Peer Var., Class Size, School F.E. Excluded
0 Years of Experience	-2.7495	-2.6024	-2.7354	-2.2279	-1.9576
	(0.91)	(0.86)	(0.90)	(0.75)	(0.66)
1 Year of Experience	1.7714	1.8920	1.7763	1.8644	2.1128
	(0.82)	(0.88)	(0.83)	(0.88)	(1.01)
2-4 Years of Experience	-0.4575	-0.4340	-0.4518	-0.4871	-0.3328
	(0.29)	(0.27)	(0.28)	(0.31)	(0.22)
Total In-service $Hours_t$	0.0115	0.0118	0.0115	0.0113	0.0116
	(1.07)	(1.10)	(1.07)	(1.06)	(1.10)
Total In-service $Hours_{t-1}$	-0.0240**	-0.0227**	-0.0241**	-0.0227**	-0.0213**
	(2.22)	(2.11)	(2.23)	(2.12)	(2.01)
Total In-service $Hours_{t-2}$	-0.0225**	-0.0217**	-0.0226**	-0.0216**	-0.0210**
	(2.06)	(2.00)	(2.08)	(2.00)	(1.96)
Total In-service $Hours_{t-3}$	-0.0155	-0.0152	-0.0157	-0.0145	-0.0150
	(1.53)	(1.50)	(1.55)	(1.45)	(1.50)
Content In-service $Hours_t$	-0.0177	-0.0171	-0.0176	-0.0185	-0.0173
	(0.92)	(0.89)	(0.92)	(0.98)	(0.92)
Content In-service $Hours_{t-1}$	0.0216	0.0190	0.0218	0.0209	0.0182
	(1.07)	(0.94)	(1.09)	(1.05)	(0.92)
Content In-service $Hours_{t-2}$	0.0517**	0.0501**	0.0521**	0.0529**	0.0516**
	(2.37)	(2.30)	(2.40)	(2.46)	(2.41)
Content In-service $Hours_{t-3}$	0.0100	0.0127	0.0101	0.0097	0.0122
	(0.42)	(0.53)	(0.42)	(0.41)	(0.52)
Advanced Degree	0.4133	0.3020	0.3965	0.4384	0.3039
	(0.24)	(0.18)	(0.23)	(0.26)	(0.18)
F-Test on Constraints		2.53**	0.90	0.79	0.83
Student Fixed Effects	Yes	Yes	Yes	Yes	Yes
Teacher Fixed Effects	Yes	Yes	Yes	Yes	Yes
School Fixed Effects	Yes	Yes	Yes	No	No
Number of Students	47,442	47,442	47,442	47,442	47,442
Number of Observations	74,196	74,196	74,196	74,196	74,196

Models include the following time varying student/class characteristics: number of schools attended by the student in the current year, "structural" move by student, "non-structural" move by student, indicator of a student repeating a grade." All models also include year, grade level, and repeater-by-grade dummies. Absolute values of t-statistics appear in parentheses. * indicates statistical significance at the .10 level and ** indicates significance at the .05 level and *** indicates significance at the .01 level in a two-tailed test.

Table 5B
Correlation of Estimated Teacher Fixed Effects From
Student Math Achievement Restricted Value-Added Models
With Differing Controls for Classroom and School Characteristics
(100 Florida Middle Schools, 1999/00-2003/04)

	Peer Var., Class Size, School F.E. Included	No. Peer Var.	No Class Size	No School F.E.	Peer Var., Class Size, School F.E. Excluded
Peer Var., Class Size, School F.E.	1.0000				
No Peer Variables	0.9540	1.0000			
No Class Size Variable	0.9134	0.9355	1.0000		
No School Fixed Effects	0.5143	0.5365	0.5050	1.0000	
No Peer, No Class Size, No School F.E.	0.5126	0.5318	0.5016	0.9880	1.0000

Table 6A
Estimates of the Impact of Teacher Characteristics on Student Math Achievement Using Restricted Value-Added Models With Differing Controls for Student Heterogeneity (100 Florida Middle Schools, 1999/00-2003/04)

Explanatory Variable	Time-Invariant Student Characteristics	Student Fixed Effects	Student Random Effects	Difference Between Fixed and Random Effects Models
0 Years of Experience	-2.8612***	-2.7346	-2.8788***	0.1442
	(2.70)	(0.90)	(2.71)	(0.07)
1 Year of Experience	-1.9354**	1.7988	-1.9571**	3.7560**
	(2.54)	(0.84)	(2.56)	(2.48)
2-4 Years of Experience	-1.8139***	-0.4381	-1.8316***	1.3936
	(3.12)	(0.28)	(3.15)	(1.28)
Total In-service $Hours_t$	0.0031	0.0116	0.0035	0.0081
	(0.75)	(1.08)	(0.85)	(1.13)
Total In-service $Hours_{t-1}$	-0.0022	-0.0241**	-0.0017	-0.0224***
	(0.55)	(2.23)	(0.44)	(3.00)
Total In-service $Hours_{t-2}$	-0.0040	-0.0226**	-0.0034	-0.0191**
	(0.99)	(2.07)	(0.86)	(2.54)
Total In-service $Hours_{t-3}$	0.0009	-0.0156	0.0013	-0.0169**
	(0.21)	(1.53)	(0.32)	(2.46)
Content In-service $Hours_t$	-0.0057	-0.0181	-0.0064	-0.0117
	(0.77)	(0.95)	(0.87)	(0.88)
Content In-service $Hours_{t-1}$	-0.0044	0.0215	-0.0048	0.0264*
	(0.58)	(1.07)	(0.64)	(1.85)
Content In-service $Hours_{t-2}$	0.0139*	0.0515**	0.0131*	0.0384**
	(1.77)	(2.36)	(1.68)	(2.43)
Content In-service $Hours_{t-3}$	-0.0050	0.0102	-0.0060	0.0161
	(0.58)	(0.42)	(0.69)	(0.96)
Advanced Degree	0.1584	0.4168	0.0575	0.3592
	(0.24)	(0.24)	(0.09)	(0.30)
Student Fixed Effects	No	No	Yes	
Teacher Fixed Effects	Yes	Yes	Yes	
School Fixed Effects	Yes	Yes	Yes	
Number of Students	47,435	47,435	47,435	
Number of Observations	74,187	74,187	74,187	

Models include the following time varying student/class characteristics: number of schools attended by the student in the current year, “structural” move by student, “non-structural move” by student, indicator of a student repeating a grade, class size, fraction of classroom peers who are female, fraction of classroom peers who are black, average age (in months) of classroom peers, fraction of classroom peers who changed schools, fraction of classroom peers who made a "structural move." All models also include year, grade level, and repeater-by-grade dummies. Absolute values of t-statistics appear in parentheses. * indicates statistical significance at the .10 level and ** indicates significance at the .05 level and *** indicates significance at the .01 level in a two-tailed test.

Table 6B
Correlation of Estimated Teacher Fixed Effects From Student Math Achievement Restricted Value-Added Models With Differing Controls for Student Heterogeneity (100 Florida Middle Schools, 1999/00-2003/04)

	Time-Invariant Student Characteristics	Student Fixed Effects	Student Random Effects
Time-Invariant Student Characteristics	1.0000		
Student Fixed Effects	0.3926	1.0000	
Student Random Effects	0.9498	0.4514	1.0000

Table 7
Estimates of the Impact of Teacher Characteristics on Student Math Achievement Using Restricted Value-Added Models – Teacher-Specific Versus Within-School Grade-Level Average Measures (100 Florida Middle Schools, 1999/00-2003/04)

Explanatory Variable	Teacher-Specific Characteristics	Within-School Grade-Level Average Teacher Characteristics
0 Years of Experience	-2.7304	-0.0499
	(0.90)	(0.01)
1 Year of Experience	1.7819	4.3829
	(0.83)	(1.23)
2-4 Years of Experience	-0.5060	-2.3871
	(0.32)	(0.71)
Total In-service $Hours_t$	0.0129	0.0598*
	(1.19)	(1.86)
Total In-service $Hours_{t-1}$	-0.0236**	-0.0107
	(2.18)	(0.35)
Total In-service $Hours_{t-2}$	-0.0230**	-0.0082
	(2.10)	(0.32)
Total In-service $Hours_{t-3}$	-0.0156	0.0146
	(1.54)	(0.63)
Content In-service $Hours_t$	-0.0185	-0.0454
	(0.96)	(0.75)
Content In-service $Hours_{t-1}$	0.0214	-0.0761
	(1.06)	(1.37)
Content In-service $Hours_{t-2}$	0.0534**	0.0251
	(2.44)	(0.38)
Content In-service $Hours_{t-3}$	0.0112	0.0679
	(0.47)	(0.87)
Advanced Degree	0.4881	1.0405
	(0.28)	(0.36)
Student Fixed Effects	Yes	Yes
Teacher Fixed Effects	Yes	No
School Fixed Effects	Yes	No
School-by-Year Fixed Effects	No	Yes
Grade-by-School Fixed Effects	No	Yes
Number of Students	47,404	47,404
Number of Observations	74,013	74,013

Models include the following time varying student/class characteristics: number of schools attended by the student in the current year, “structural” move by student, “non-structural move” by student, indicator of a student repeating a grade, class size, fraction of classroom peers who are female, fraction of classroom peers who are black, average age (in months) of classroom peers, fraction of classroom peers who changed schools, fraction of classroom peers who made a "structural move." All models also include repeater-by-grade dummies. Absolute values of t-statistics appear in parentheses. * indicates statistical significance at the .10 level and ** indicates significance at the .05 level and *** indicates significance at the .01 level in a two-tailed test.