3-1-2012

SKILLS, PRODUCTIVITY AND THE EVALUATION OF TEACHER PERFORMANCE

Douglas N. Harris
University of Wisconsin – Madison, dnharris3@wisc.edu

Tim R. Sass
Georgia State University, tsass@gsu.edu

W.J. Usery Workplace Research Group Paper Series

Working Paper 2012-3-1
March 2012

Skills, Productivity and the Evaluation of Teacher Performance

Douglas N. Harris
University of Wisconsin - Madison

Tim R. Sass
Georgia State University

Georgia State University
Andrew Young School of Policy Studies

SKILLS, PRODUCTIVITY AND

THE EVALUATION OF TEACHER PERFORMANCE *

by

Douglas N. Harris
Department of Educational Policy Studies
La Follette School of Public Affairs
University of Wisconsin – Madison
dnharris3@wisc.edu

Tim R. Sass
Department of Economics
Andrew Young School of Policy Studies
Georgia State University
tsass@gsu.edu

March 1, 2012

Abstract

We examine the measurement and prediction of worker productivity using a sample of teachers and school principals. We find that principals' evaluations are positively associated with teachers' estimated contributions to students' test scores (value-added), and are better predictors of teacher value-added than are teacher credentials. Principals' assessments of teachers' cognitive and non-cognitive skills are strongly associated with principals' overall teacher evaluations and to a lesser extent with teacher value-added. While past teacher value-added predicts future value-added, principals' subjective ratings can provide additional information, particularly when prior value-added measures are based on a single year of teacher performance.

* This study is funded under grant R305M040121 from the U.S. Department of Education. We wish to thank Stacey Rutledge, William Ingle, Peter Hinrichs and participants in the NBER summer 2007 education workshop for their comments. Julia Manzella, John Gibson, William Ingle, Micah Sanders, Cynthia Thompson and Lisa Verdon provided valuable research assistance. Previous versions of the paper circulated under the title "What Makes for a Good Teacher and Who Can Tell?" All remaining errors are the responsibility of the authors.

I. Introduction

Recent research consistently finds that teacher productivity is the most important component of a school's effect on student learning and that there is considerable heterogeneity in teacher productivity within and across schools (Rockoff (2004), Hanushek, et al. (2005), Rivkin, Hanushek and Kain (2005), Kane, Rockoff and Staiger (2008), Aaronson, Barrow and Sander (2007)). However, relatively little is known about what makes some teachers more productive than others in promoting student achievement. The first few years of teacher experience improve productivity (Rockoff (2004), Hanushek, et al. (2005), Jepsen (2005), Rivkin, Hanushek and Kain (2005), Boyd, et al. (2006), Clotfelter, Ladd and Vigdor (2006, 2007, 2010), Kane, Rockoff and Staiger (2008), Harris and Sass (2011)). But little else in the way of observed teacher characteristics seems to consistently matter.[1] Thus, while teachers significantly impact student achievement, the variation in teacher productivity is still largely unexplained by commonly measured teacher characteristics.

One possible explanation for the inability of extant research to identify the determinants of teacher productivity is that researchers have not been measuring the characteristics that truly affect productivity. Recent work in labor economics suggests, for example, that personality traits such as conscientiousness play an important role in determining worker productivity (Borghans, ter Weel, and Weinberg (2008), Cunha, Heckman, Lochner, and Masterov (2006); Heckman, Stixrud, and Urzua (2006)). But the relative predictive value of cognitive and

---

[1] Harris and Sass (2011) find that the gains to experience may extend well beyond the first few years. Clotfelter, Ladd and Vigdor (2007, 2010), using North Carolina data, find some teacher credentials are correlated with teacher effectiveness, particularly at the secondary level. Goldhaber (2007) also uses the North Carolina data and finds similar results, though he questions the signal value of credentials that are weakly correlated with productivity.

non-cognitive factors is hard to assess due to the difficulty in obtaining measures of both cognitive and non-cognitive skills and labor productivity.

Unraveling the factors associated with teacher productivity could yield valuable insights into the most appropriate policies for selecting and training teachers. If teacher productivity is affected primarily by personality characteristics that are measurable ex-ante, they could be used to screen applicants and identify the most desired candidates in the hiring process. If, however, the most important teacher characteristics are malleable, such as subject content knowledge, determining which teacher characteristics have the greatest impact on student learning could inform the design of pre-service and in-service teacher training programs

Intertwined with the relationship between teacher characteristics and teacher productivity is the issue of how best to evaluate teacher performance. If teacher productivity is not strongly correlated with teacher credentials like educational attainment, but is associated with skills that can be discerned through observing behavior, direct monitoring and evaluation of teacher performance may be advantageous. Consistent with this idea, school principals are being granted greater authority in hiring, evaluation and retention of teachers both through the creation of independent charter schools nationwide and through decentralization reforms in public school districts such as New York City. Spurred on by the federal *Teacher Incentive Fund* (TIF) and *Race to the Top* (RTTT) initiatives, classroom observations by principals or external evaluators are increasingly being used to make high-stakes decisions about employment and compensation.[2] The downside of subjective evaluations by principals is they may be affected by

---

[2] For in-depth discussions of performance-based compensation in schools see Figlio and Kenny (2007) and Podgursky and Springer (2007).

personal bias toward factors unrelated to productivity and some principals may simply be poor judges of teacher productivity.

Another possibility is that neither teacher credentials nor observable traits are correlated with a teacher's productivity. In this case ex-post evaluation of teachers based on their contributions to student achievement, or "value-added," may be optimal (Gordon, Kane, and Staiger (2006)). The TIF and RTTT reforms also generally involve teacher value-added as part of the overall evaluation. However, there are concerns about the precision of value-added measures, their narrow focus on student test scores, and the fact that they can only be calculated for a small proportion of teachers (Baker et al. (2010); Harris (2011)).

In this paper we consider the two interrelated issues of the skills associated with teacher productivity and how best to evaluate teacher performance. Specifically, we address the following four questions:

1) What characteristics of teachers are associated with teacher productivity?
2) What information do principals use in assessing teacher performance?
3) How closely are principal evaluations associated with teacher productivity?
4) How well do principal evaluations and prior measures of teacher productivity predict *future* teacher productivity?

We build on extant research in four ways. First, we go beyond general ratings of teacher ability and estimate the relationship between a variety of specific teacher personality traits and teacher productivity. Second, we test how well prior value-added scores and prior principal evaluations of teachers predict *future* teacher value-added. The ability to predict future performance is critical, especially when probationary employment periods are followed by decisions that provide long-term job security to employees (e.g., teacher tenure) that can affect organizational performance for years or decades. Third, unlike other existing studies, we consider how the relationships between teacher characteristics, principal evaluations and teacher value-added vary between elementary and middle/high schools, each of which has distinctive organizational structures. Finally, while most of our results are from value-added measures based on a low-stakes test, we also consider results from high-stakes achievement tests, though the usefulness of these is limited by much smaller samples in our data set.

In the next section we describe the small existing literature on subjective evaluations of teachers and their relationship with value-added. This is followed by a discussion of the data used for our analysis, including how the interviews with principals were conducted and our method for estimating teacher value-added. In the concluding section we discuss our empirical results and possible policy implications.

II. Literature Review

The labor economics literature increasingly integrates theories and research from psychology. For example, Cuhna, et al. (2006) model the life cycle of skill attainment, giving a prominent position to personality traits. Borghans, ter Weel, and Weinberg (2008) theorize that different types of jobs require different combinations of personality traits, especially "directness" and "caring," and find evidence that some of these traits are correlated with productivity. This is perhaps not surprising, especially for jobs (such as teaching) that require substantial interpersonal interaction and communication, but it does suggest that economists may need to consider more than intelligence when evaluating the role of innate ability in labor market outcomes (Borghans, et al. (2008)).

Personality traits are difficult to measure objectively (Borghans, et al. (2008)) and perhaps more easily captured through direct observation. For this reason, the importance of personality traits in determining productivity may also affect the optimal mix of subjective supervisor ratings and more objective measures of output in evaluating and compensating workers. There is a long history of research studying the relationships between subjective and objective measures of worker productivity, as well as the implications of this relationship for optimal employment contracts. As noted by Jacob and Lefgren (2008), this research suggests that there is a relatively weak relationship between subjective and objective measures (Bommer, et al. (1995), Heneman (1986)). One reason might be that supervisors are heavily influenced by personality traits, more so than is warranted by the role personality actually plays in (objective) productivity. This interpretation is reinforced by evidence that evaluators' subjective assessments are biased, in the sense that certain types of workers (e.g., females and older workers) receive lower subjective evaluations for reasons that appear unrelated to their actual productivity (e.g., Varma and Stroh (2001)).

There is a limited literature that specifically addresses the relationship between subjective and objective assessments of school teachers. Subjective evaluations by school principals are especially interesting because principals are typically required to observe teachers and they collect a lot of information informally, and inexpensively, in the natural course of being in the school, interacting with teachers, talking with parents and so on. Three older studies have examined the relationship between student test scores and principals' subjective assessments using longitudinal student achievement data to measure student learning growth (Murnane (1975), Armor, et al. (1976), and Medley and Coker (1987)). However, as noted by Jacob and Lefgren (2008), these studies do not account for measurement error in the objective test-based measure and therefore under-state the relationship between subjective and objective measures.

In their work, Jacob and Lefgren address both the selection bias and measurement error problems within the context of a value-added model for measuring teacher productivity that is linked to principals' subjective assessments. They obtain student achievement data and combine it with data on principals' ratings of 201 teachers in a mid-sized school district in a Western state.[3] Jacob and Lefgren find that previous teacher value-added is a better predictor of current student achievement than are current principal ratings. In particular, teacher value-added calculated from test scores in 1998-2002 was a significantly better predictor of 2003 test scores (conditional on student and peer characteristics) than were 2003 principal ratings made just prior to the 2003 student exam. The current principal ratings were also significantly correlated with current test scores, conditional on prior value-added. While this latter finding suggests contemporaneous principal ratings add information, the reason is not clear. The principal ratings might provide more precise indicators of previous teacher productivity (especially when there is little prior test score information, as is often the case), since past value-added is subject to transient shocks to student test scores. Alternatively, the principal ratings may simply reflect new current-school-year (2002/03) performance information not included in past value-added (based on test scores through 2001/02). In order to sort out these effects, in our analysis we compare the ability of current value-added and current principal ratings to predict *future* teacher value-added.

The only prior study to consider principals' assessments of specific teacher characteristics, as opposed to the overall rating, is an unpublished working paper by Jacob and Lefgren (2005). While they find a positive and significant relationship between the teacher value-added and teachers' relationship with the school administration, this is the only teacher characteristic they consider.

Rockoff, et al. (forthcoming) study an experiment in which elementary and middle school principals in New York City were randomly assigned to receive teacher value-added information. They found that principals change their evaluations of teachers when they receive new information about the impact of

[3] As in the present study, the district studied by Jacob and Lefgren chose to remain anonymous.

teachers on student test scores. The extent of updating is positively related to the precision of value-added information they receive and negatively related to the quality of their own prior information on teachers. The acquisition of new information also appears to have significant effects on personnel decisions and student outcomes. Rockoff, et al. find that teachers with low value-added scores were more likely to exit their schools after the principal received value-added information which in turn led to a small increase in student test scores. While not the focus of their analysis, Rockoff, et al. also estimate pre-experiment correlations between various value-added measures and principals' evaluations of their teachers. They find positive correlations, similar in magnitude to those obtained by Jacob and Lefgren. The correlations tend to increase with the precision of the value-added estimates and with the number of years the principal has known a teacher.

Most recently, Kane, et al. (2012) report interim results from the Measures of Effective Teaching (MET) project, sponsored by the Gates Foundation. The project measures teacher productivity in different ways, including through student evaluations, observations of classroom practice by trained evaluators, and, as in our study, student performance on two different achievement tests.[4] The MET project does not include principal evaluations, however. Kane, et al. compare the ability of teacher observations, student surveys and value-added in one classroom

---

[4] A number of other studies have examined the relationship between the achievement levels of teachers' students and subjective teacher ratings that are based on formal standards and extensive classroom observation (Gallagher (2004), Kimball et al. (2004), Milanowski (2004)). For example, in Milanowski (2004), the subjective evaluations are based on an extensive standards-framework that required principals and assistant principals to observe each teacher six times in total and, in each case, to rate the teacher on 22 separate dimensions. All of these studies find a positive and significant relationship, despite differences in the way they measure teacher value-added and in the degree to which the observations are used for high-stakes personnel decisions. While these studies have the advantage of more structured subjective evaluations, the reliance on achievement levels with no controls for lagged achievement or prior educational inputs makes it difficult to estimate teacher value-added.

to predict teacher value added in another course section by the same teacher in the same time period.[5] They find that the best predictor is value added, with only a modest improvement from adding in classroom observations and student feedback.

III. Data and Methods

We begin by describing the general characteristics of the school district and sample of principals, teachers and students. We then discuss in more detail the two main components of the data: (a) administrative data that are used to estimate teacher value-added; and (b) principal interview data that provide information about principals' overall assessments of teachers as well as ratings of specific teacher characteristics.

A. General Sample Description

The analysis is based on interviews with 30 principals from an anonymous mid-sized Florida school district. The district includes a heterogeneous population of students. For example, among the sampled schools, the school-average proportion of students eligible for free/reduced price lunches varies from less than 10 percent to more than 90 percent. Similarly, there is considerable heterogeneity among schools in the racial/ethnic distribution of their students. We interviewed principals from 17 elementary (or K-8) schools, six middle schools, four high schools, and three special population schools, representing more than half of the principals in the district. The racial distribution of interviewed principals is comparable to the national average of all principals (sample district: 78 percent White; national: 82 percent White) as is the

[5] They also evaluate the relative performance of the measures to "predict" past performance by the same teacher.

percentage with at least a master's degree (sample district: 100 percent; national: 90.7 percent).[6] However, the percentage female is somewhat larger (sample district: 63 percent; national: 44 percent).

The advantage of studying a school district in Florida is that the state has a long tradition of strong test-based accountability (Harris, Herrington and Albee, 2007) that has now come to pass in other states as a result of the federal *No Child Left Behind* policy. The state has long graded schools on an A-F scale. The number of schools receiving the highest grade has risen over time; in our sample 20 schools received the highest grade (A) during the 2005-06 school year; the lowest performing school in the district received a grade of D. It is reasonable to expect that accountability policies, such as the school grades mentioned above, influence the objectives that principals see for their schools and therefore their subjective evaluations of teachers. For example, we might expect a closer relationship between value-added and subjective assessments in high accountability contexts where principals are not only more aware of test scores in general, but where principals are increasingly likely to know the test scores, and perhaps test score gains, made by students of individual teachers. We discuss the potential influence of this phenomenon later in the analysis, but emphasize here that, by studying a Florida school district, the results of our analysis are more applicable to the current policy environment where high-stakes achievement-focused accountability is federal policy.

[6] The national data on principals comes from the 2003-2004 Schools and Staffing Survey (SASS) as reported in the Digest of Education Statistics (National Center for Education Statistics, 2006). Part of the reason that this sample of principals has higher levels of educational attainment is that Florida law makes it difficult to become a principal without a master's degree.

B. Student Achievement Data and Modeling

Throughout Florida there is annual testing in grades 3-10 for both math and reading. Until recently, two tests were administered, a high-stakes, criterion-referenced exam based on the state curriculum standards known as the FCAT-Sunshine State Standards (SSS) exam, and a low-stakes, norm-referenced test (NRT) which is the Stanford Achievement Test. We mainly employ the low-stakes NRT in the present analysis for two reasons. First, it is a vertically scaled test, meaning that unit changes in the achievement score should have the same meaning at all points along the scale. Second, and most importantly, the district under study also administers the NRT in grades 1 and 2, allowing us to compute achievement gains for students in grades 2-10. Achievement data on the NRT are available for each of the school years 1999/00 through 2007/08.[7] The SSS exam was instituted a year later and thus scores on the high-stakes test are only available for the 2000/01-2007/08 school years. Using the low-stakes test we are able to estimate the determinants of achievement gains for five years prior to the principal interviews, 2000/01-2005/06, and for two years after the interviews, 2006/07-2007/08. In order to account for any differences in test content and scaling across grades across time, we normalize test scores by grade and year. Characteristics of the sample used in the value-added analysis are described in Table 1.

In order to compute value-added scores for teachers we estimate a model of student achievement, A_t, of the following form:

$$A_{it} = \lambda A_{it-1} + \boldsymbol{\beta}_1 \mathbf{X}_{it} + \boldsymbol{\beta}_2 \mathbf{P}_{-ijmt} + \delta_k + \phi_m + \rho_g + \gamma_{gt} + \nu_{it} \quad (1)$$

[7] Prior to 2004/05 version 9 of the Stanford Achievement Test (SAT-9) was administered. Beginning in 2004/05 the SAT-10 was given. All SAT-10 scores have been converted to SAT-9 equivalent scores based on the conversion tables in Harcourt (2002).

The effects of prior educational inputs are captured by the lagged test score, A_{it-1}, and are assumed to diminish geometrically over time at a rate (1-λ). The vector $\mathbf{X}_{it}$ includes time-varying student characteristics such as student mobility, free/reduced-price lunch eligibility and limited English proficiency status as well as time-constant student attributes like race/ethnicity and gender. The vector of peer characteristics, $\mathbf{P}_{-ijmt}$ (where the subscript –i students other than individual i in the classroom), includes both exogenous peer characteristics and the number of peers or class size. In addition, a teacher fixed effect (δ_k), a school fixed effect ($\phi_{m)}$ and a sets of grade-repeater-by-grade (ρ_g) and grade-by-year indicators (γ_{gt}) are also included.[8] The teacher fixed effect captures both the time-invariant characteristics of teachers as well as the average value of time-varying characteristics like experience and possession of an advanced degree. Since school fixed effects are included, the estimated teacher effects represent the "value-added" of an individual teacher relative to the average teacher at the school. The final term, ν_{it}, is a mean zero random error.

The achievement model depicted in equation (1) is but one of many commonly estimated value-added models. We utilize it as our primary model since recent experimental and simulation-based evidence suggests it is likely to produce relatively unbiased estimates of teacher effects under a range of conditions ((Kane and Staiger (2008), Guarino, Reckase and Wooldridge (2011)). However, we show in Appendix Table A2 that the relationship between value-added estimates of teacher productivity and principal evaluations of teacher quality are similar when other value-added models, ones that assume complete

[8] The full set of student and peer explanatory variables is delineated in appendix table A1.

persistence in prior inputs or control for student heterogeneity with student fixed effects, are employed.[9]

Recently, Rothstein (2010) has argued that value-added models may produce biased estimates of teacher productivity due to the non-random assignment of students to teachers within schools. For example, if students who experience an unusually high achievement gain in one year are assigned to particular teachers the following year and there is mean reversion in student test scores, the estimated value-added for the teachers with high prior-year gains will be biased downward. Rothstein proposes falsification tests based on the idea that future teachers cannot have causal effects on current achievement gains. We conduct falsification tests of this sort, using the methodology employed by Koedel and Betts (2011). For each level of schooling, elementary, middle and high, we fail to reject the null of strict exogeneity, indicating that the data from the district we analyze in this study are not subject to the sort of dynamic sorting bias concerns raised by Rothstein.[10]

As noted by Jacob and Lefgren, another concern is measurement error in the estimated teacher effects. Given the variability in student test scores, value-added estimates will yield "noisy" measures of teacher productivity, particularly for teachers with relatively few students (McCaffrey, et al (2009)). We employ three strategies to alleviate the measurement error problem. First, we limit our sample to teachers who taught at least five students with achievement gain data. Second, we employ the measurement-error correction procedure adopted by Jacob and Lefgren when evaluating the strength of correlations between value-added

---

[9] For a thorough discussion of various value-added models and the assumptions that underlie them, see Todd and Wolpin (2003) and Harris, Sass and Semykina (2011).

[10] The failure to reject strict exogeneity may indicate that prior test-score gains are not used to assign students to teachers in the studied district or to the fact that we are including many cohorts of students per teacher in our analysis (Koedel and Betts (2011)).

and subjective evaluations by principals.[11] Third, in regression analyses where value-added is the dependent variable we use a feasible generalized least squares (FGLS) estimation procedure which accounts for estimation error in the dependent variable.[12] As noted by Mihaly et al. (2010), standard fixed-effects software routines compute fixed effects relative to some arbitrary hold-out unit (e.g. an omitted teacher), which can produce wildly incorrect standard errors and thus inappropriate corrections for measurement error in the estimated teacher effects. Therefore, to estimate the teacher effects and their standard errors we employ the Stata routine *felsdvregdm*, developed by Mihaly et al. (2010), which imposes a sum-to-zero constraint on the teacher estimated teacher effects within a school and produces the appropriate standard errors for making measurement error adjustments.[13]

C. Principal Interview Data

Interviews were conducted in the summer of 2006. Each principal was asked to rate up to ten teachers in grades and subjects that are subject to annual student achievement testing. Per the requirements of the district, the interviews were "single-blind" so that the principal knew the names of the teachers but the interviewer knew only a randomly assigned number associated with the names.

From the administrative data described above, we identified teachers in tested grades and subjects in the 30 schools who had taught at least one course with 10 or more tested students and who were still in the school in the 2004/05 school year (the last year for which complete administrative data were available

---

[11] See Jacob and Lefgren (2008), p.113.

[12] Specifically, we employ the method developed by Lewis and Linzer (2005) and embodied in the Stata routine *edvreg*.

[13] All standard errors on the estimated teacher effects are corrected for clustering at the classroom level using the method suggested by Moulton (1990).

prior to arranging the principal interviews). In some cases, there were fewer than ten teachers who met these requirements. Even in schools that had ten teachers on the list, there were cases where some teachers were not actually working in the respective schools at the time of the interview. If the principal was familiar with a departed teacher and felt comfortable making an assessment, then these teachers and subjective assessments were included in the analysis. If the principal was not sufficiently familiar with the departed teacher, then the teacher was dropped. Many schools had more than ten teachers. In these cases, we attempted to create an even mix of five teachers of reading and math. If there were more than five teachers in a specific subject, we chose a random sample of five to be included in the list.

In the interviews, principals were first asked to mark on a sheet of paper the principal's overall assessment of each teacher, using a 1-9 scale.[14] The interviewer then handed the principal another sheet of paper so that he/she could rate each teacher on each of 12 characteristics: caring, communication skills, enthusiasm, intelligence, knowledge of subject, strong teaching skills, motivation, works well with grade team/department, works well with me (the principal), contributes to school activities beyond the classroom, and contributes to overall school community. The first seven characteristics in this list were found by Harris, Rutledge, Ingle, and Thompson (2010) to be among the most important characteristics that principals look for when hiring teachers.[15] Having an

[14] The specific question was: "First, I would like you to rate each of the ten teachers relative to the other teachers on the list. Please rate each teacher on a scale from 1-9 with 1 being not effective to 9 being exceptional. Place an X in the box to indicate your choice. Also please circle the number of any teachers whose students are primarily special populations."

[15] As described in Harris, Rutledge, Ingle and Thompson (2010), the data in this study came from the second in a series of interviews carried out by the researchers. During the summer of 2005, interviews were conducted regarding the hiring process and principals preferred characteristics of teachers. The first set of interviews was important because it helped validate the types of teacher characteristics we consider. Principals were asked an open-ended question in the first interview

occupation-specific list of characteristics is important because recent economic theory and evidence suggest that different traits matter more in different occupations and specifically that "caring" is more important in teaching than in any other occupation (Borghans, ter Weel, and Weinberg (2008)).

The interview questions were designed so that principals would evaluate teachers relative to others in the school.[16] One reason for doing so is that even an "absolute" evaluation would be necessarily based on each principal's own experiences. This implies that ratings on individual characteristics across principals may not be based on a common reference point or a common scale. Therefore, like Jacob and Lefgren, we normalize the ratings of each teacher characteristic to have a mean of zero and standard deviation of one over all teachers rated by a given principal. Given our teacher fixed-effects estimates are within-school measures, normalizing the ratings allow us to compare within-school ratings to within-school teacher value-added.

The final activity of the interview involved asking the principals to rate each teacher according to the following additional "outcome" measures: raises FCAT math achievement, raises FCAT reading achievement, and raises FCAT writing achievement. These last measures help us test whether the differences between the value-added measures and the principals' overall assessments are due to philosophical differences regarding the importance of student achievement as

about the teacher characteristics they prefer. Two-thirds of these responses could be placed in one of 12 categories identified from previous studies on teacher quality. The list here takes those ranked highest by principals in the first interview and then adds some of those included by Jacob and Lefgren.

[16] In contrast, in the Rockoff, et al. (forthcoming) study, principals were asked to compare each teacher to all "teachers [they] have known who taught the same grade/subject," not just teachers at their own school.

an educational outcome or to difficulty in identifying teachers who increase student test scores.

To lessen potential multicollinearity problems and reduce the number of teacher characteristics to analyze, we conduct a factor analysis of the 11 individual teacher characteristics rated by principals. As indicated in Table 2, the individual characteristics can be summarized into four factors: interpersonal skills, motivation/enthusiasm, ability to work with others, and knowledge/teaching skills/intelligence.

Finally, as part of the interview, we discovered that principals have access to a district-purchased software program, Snapshot^{TM}, that allows them to create various cross-tabulations of student achievement data on the high-stakes SSS exam, including simple student learning gains and mean learning gains by teacher. While we have no data about the actual usage of this software, subsequent informal conversations with two principals suggests that at least some principals use the program to look at the achievement gains made by students of each teacher. While this may have provided principals with some information about unconditional student average achievement gains, that is of course not the same thing as the teacher value-added scores, which are conditional on student and peer characteristics. Nevertheless, we calculate and analyze alternative teacher performance measures that approximate those that principals may have seen with the software package.

IV. Results

In order to compute value-added scores for teachers we estimate equation (1) using data on current and lagged test scores for grades 2-10 over the period 2000/01 through 2005/06 with the low-stakes NRT test and report these results throughout the main text.

A. The Association between Teacher Traits and Teacher Productivity

Simple pairwise correlations among the estimated teacher fixed effects and the four teacher characteristic factors are presented in Table 3. For correlations with value-added we include correlations adjusted for estimation error in the teacher effects. There are positive relationships between teacher value-added in math and each of the teacher characteristic factors; correlations adjusted for estimation error are each in the range of 0.19 to 0.34. For reading, the adjusted correlations between teacher traits and teacher productivity in promoting student achievement are in the range of 0.20 to 0.45.

Teacher personality traits are all positively and strongly correlated with one another in both subjects; correlations are in the range of 0.61 to 0.76. It is not obvious that this should be the case, e.g., that teachers who are more knowledgeable would also tend to have better interpersonal skills. There might be a “halo effect” whereby teachers who are rated highly by the principal overall are automatically given high marks on all of the individual characteristics, though this is very difficult to test without having some other independent measure of teacher characteristics. Finally, note that among the four teacher characteristic factors, knowledge/teaching skills/intelligence is most closely associated with teacher value-added in math while value-added estimates of teacher performance in reading are most closely associated with motivation/enthusiasm of the teacher.[17]

---

[17] Our comparisons in Table 3 are between value-added based on all prior available information and a single-year principal evaluation. If teacher productivity varies over time and principals weight current performance more than the correlations could differ if we instead had used only recent achievement data to form the value-added measures of teacher productivity. We explore this issue in appendix table A3, where we present correlations of value added with principal ratings and teacher traits using varying time periods to calculate value added. We find that the correlations with contemporaneous (2005/06) value added are quite similar to those with value added constructed from all prior information (1999/00-2005/06).

A multivariate analysis of the relationship between teacher traits and teacher value-added is presented in Table 4. Results in column [1] indicate that knowledge/teaching skills/intelligence is positively and significantly associated with teacher value-added in math only. None of the other coefficients in column [1] are significant. Column [2] shows that the magnitude of the effect on math achievement of knowledge/teaching/intelligence skills is nearly identical in elementary and middle/high school, though the precision is much lower in middle/high school. The overall explanatory power of the four factors is quite low, however, with R-squared values of 0.14.[18] For reading, the only factor which is statistically significant is teacher motivation/enthusiasm. Once again, the effect is equal in magnitude across grades, but only statistically significant for elementary school teachers. The relative importance of subject matter knowledge in math teacher performance is consistent with recent findings that "Teach for America" teachers, who possess exceptionally strong academic credentials, tend to outperform traditionally prepared teachers in teaching math, but are on par with traditionally prepared teachers in reading instruction.[19]

B. The Information Principals Use in Forming Their Teacher Assessments

To determine what factors are important to a principal in forming their evaluation of a teacher we compute pairwise correlations among the four teacher characteristic factors and two measures of a principal's rating of teachers: the overall rating and the rating of the teacher's ability to raise test scores. The results, presented in Table 5, indicate that the overall rating is highly correlated

---

[18] Some of the insignificant effects may be due to multicollinearity. As demonstrated in Table 3, the four factors are all positively correlated. When each factor is regressed on estimated teacher effects separately, all are significant except "works well with others" in predicting the value-added of reading teachers.

[19] See Boyd, et al. (2006), Kane, Rockoff and Staiger (2008) and Xu, Hannaway and Taylor (2011).

with the "ability to raise test scores" rating (0.73 in math and 0.74 in reading). As with value added, principal ratings are most highly correlated with knowledge/teaching skills/intelligence in math and motivation/enthusiasm in reading.

A multivariate analysis of the relationship between the teacher characteristic factors and overall principal ratings is presented in Table 6. For math, knowledge/teaching skills/intelligence contributes the most to the principals' overall rating at all grade levels. At the elementary level, "works well with others" and interpersonal skill are also statistically significant, but the point estimates are much smaller. In reading, motivation/enthusiasm and interpersonal skill are each statistically significant across all grade levels while knowledge/teaching skills/intelligence is only significantly different from zero at the elementary level. Across both subject the four factors explain more than 80 percent of the variation in overall ratings, suggesting that the underlying 12 characteristics are important determinants of principals' overall ratings.[20]

We noted above how principal ratings may be influenced by knowledge of student test scores, especially since at least some interviewed principals had access to software that provides student average test score gains broken down by teacher. The achievement gain data accessible to principals are based on the developmental scale score (DSS) derived from the high-stakes SSS exam. In Table 7 we show that principals' overall rankings of teachers are positively correlated with once-lagged average SSS gains for math teachers, but overall rankings of reading teachers are not significantly correlated with average student achievement gains. The "ability to raise test scores" ratings are uncorrelated with

[20] This result could also be driven by the halo effect described earlier. However, principals were also asked to describe their teachers in their own words. In a working paper (Harris, Ingle, and Rutledge, 2012), we have found that the vast majority of their responses could be placed in one of the 12 categories, suggesting that these categories are indeed the main ones principals think about.

average student test-score gains in both math and reading, suggesting that principals did not use the computer software to evaluate teacher performance. Thus while recent evidence produced by Rockoff et al in New York suggests that principals incorporate value-added information in their assessment of teachers, in our Florida example we do not find strong evidence that principals are influenced by having access to simple (e.g., not conditioned on student characteristics) average student test-score gains.

C. The Association between Principal Evaluations and Teacher Value-Added

Table 8 presents FGLS estimates of the determinants of the teacher fixed effects, which account for estimation error in the teacher effects. The first column reports estimates where only standard teacher credentials (experience, possession of an advanced degree) are included as explanatory variables. None of the credential variables is a statistically significant determinant of teacher value-added scores.[21] .

In contrast, when a principal's overall rating of a teacher or their assessment of a teacher's ability to raise test scores is added to the model, its coefficient is positive and highly significant for both reading and math. (The coefficients on teacher credentials are largely unchanged.) This suggests that principals have knowledge about teacher productivity that is not captured by the standard measures of experience and educational attainment that typically form the basis for teacher pay scales. The magnitudes of the coefficients are

---

[21] In another study using statewide data from Florida (Harris and Sass (2011)), the effects of teacher experience are highly significant when teacher fixed effects are excluded, but within-teacher changes in experience are less often statistically significant. The finding that experience is insignificant in models with teacher fixed effects could mean that some apparent cross-teacher experience effects are due to attrition of less effective teachers early in their careers or that there is simply insufficient within-teacher variation in experience over a short panel The lack of significance may also be due to the relatively small sample size and the fact that the district being studied has a relatively high average level of teacher experience.

substantial. For example, the coefficient on principals' overall ratings for math teachers in Table 8 is 0.059, which implies that a teacher who is rated one point higher on the 1-9 scale raises student math test scores six hundreths of a standard deviation. Put differently, given the standard deviation in principal ratings is 1.68, a one-standard deviation in increase in the principal's overall rating of a math teacher corresponds to a 0.10 standard deviation increase in the teacher's student's test scores or moving students from the 50th to the 54th percentile.

In Table 9 we present estimates where the correlation between principal ratings and estimated teacher value-added is allowed to vary between elementary school and middle/high school. At both the elementary and middle/high school levels, the overall principal rating is positively and statistically significantly associated with the teacher fixed effect in both reading and in math. The "ability to raise test scores" rating is also statistically significant in all but middle/high school reading.[22] However, the association between teacher value-added and a one-point increase in the principal's rating scale on teacher value-added in reading is generally smaller than for math. This is consistent with the general finding in the literature that the effects of teacher characteristics on student achievement tend to be less pronounced in reading. It is often suggested that reading scores are more likely to be influenced by factors outside of school; students may read books in their free time, but they seldom work math problems for enjoyment. Alternatively, principals may not be as good at evaluating the performance of teachers in reading instruction.

---

[22] Three of the statistically significant coefficients in Table 10 (overall ratings in middle/high math and reading; ability to raise test scores in middle/high math) are statistically insignificant in the model that includes student fixed effects. Put differently, there are no statistically significant partial correlations between any of the middle high school principal ratings and teacher value-added. See appendix tables for details.

One would expect that the longer a principal has known their teachers the more accurate would be the principal's evaluation of their performance. Further, principals may gain general human capital in personnel evaluation as their experience as a supervisor increases. To test these ideas we regress the correlation between teacher fixed effects and principal evaluations on the duration of a principal's tenure as principal at their current school (representing teacher-specific knowledge) and the principal's years of experience in educational administration (a proxy for general personnel evaluation skills). Results are presented in Table 10. We find mixed evidence regarding the above hypotheses. School-specific experience and general administrative experience are not significantly correlated with the relationship between value-added and principal ratings of reading teachers. With small sample sizes of 22-26 principals, the limited statistical significance is unsurprising.

The coefficient on principals' tenure in the school is positive and significant in the regression predicting the correlation between value-added and a principal's overall evaluation of math teachers, but the opposite holds when we turn to principal school-specific tenure. The negative coefficient on "ability to raise test scores" is surprising, but there are at least two possible explanations. Principals, as they get to know teachers, may begin to assess them more based on their relationships (i.e., how well the principal gets along teachers) rather than objective performance, though this hypothesis is difficult to test with these data. Also, we estimated the value-added models with a second specification (the same as equation (1) but with the addition of student fixed effects) and the negative coefficient is essentially zero in that case (see appendix tables).

In addition to general correlations, we also consider the ability of principals to identify productive teachers at various parts of the teacher value-added distribution. In Table 11 we present cross-tabulations of the rankings of teacher value-added and principals' ratings of teachers on both the "overall" and "ability to raise test scores" metrics.[23] It appears that for both math and reading, principals are better at identifying low value-added teachers, rather than top-performing teachers. Of those teachers who rank in the bottom 30 percent based on value-added in math, 65 percent are also ranked in the bottom 30 percent by their principal. In contrast, only 16 percent of teachers in the top 30 percent in math value-added are also ranked in the 30 percent by their principal. Similar differences appear for reading teachers.[24]

Our findings differ from those of Jacob and Lefgren (2008). They find that principals are relatively good at distinguishing both high and low value-added teachers, with a somewhat better alignment of principal ratings with teacher value-added for the top rated teachers. At least two possible explanations for the divergent results come to mind. The underlying distribution of teacher quality may be more uniform in our sample, making it more difficult to distinguish the best teachers from middling teachers. Alternatively, some groups of principals may simply be better than others in identifying high-value-added teachers. In either case, the principals in both Jacob and Lefgren's analysis and in the present study seem to be able to identify their lowest-performing teachers.

D. The Relative Ability of Prior Performance and Teacher Rankings to Predict Future Teacher Performance

To this point, as in all prior studies, we have been comparing principal evaluations of teachers with value-added measures constructed from all available prior student test scores (i.e. principal ratings from summer 2006 with value-

[23] In addition to the three-category rankings presented in Table 5, we also computed cross-tabulations based on quintile rankings. The patterns of results were very similar. Given there are at most 10 teachers per school, the three-category ranking system seems more appropriate.

[24] The fact that in the high (low) category, teachers can only move down (up) whereas those in the middle can move in two directions could explain why the proportion of teachers with similar rankings is higher in the middle category than in the top category, but it does not explain why the proportion with similar rankings in the bottom category is higher.

added based on achievement data up through the 2005/06 school year). Such contemporaneous estimates of teacher productivity are relevant to decisions about the role of principal evaluations in measuring and rewarding past performance. However, contemporaneous measures of teacher performance are not particularly relevant for retention and tenure decisions, where the decision should (optimally) be based on predictions about future performance.

We measure future teacher productivity by re-estimating equation (1), using data on student achievement gains from the 2006/07 and 2007/08 school years (including test scores from 2005/06 as the initial lagged value) to derive estimates of future teacher value-added. As demonstrated by (McCaffrey, et al. (2009)), basing teacher value-added on two years of performance leads to much more precise estimates than relying on a single estimated test score gain, as in Jacob and Lefgren (2008). We then regress our estimate of future value-added on either the principal's overall rating of the teacher from the summer of 2006 or the estimated teacher fixed effect from a student achievement model covering the years 1999/00-2005/06.[25]

As shown in Table 12, we estimate the equation several ways, varying the amount of information used to estimate the past teacher value-added. With the exception of the value-added measure constructed from only 2005/06 data, we utilize a common sample to ensure comparability of the value-added estimates. The sample of teachers with value-added data for 2005/06 is much smaller, however. This is because we selected teachers to participate in the study in the

---

[25] In addition to the estimates reported in Table 12, we also estimated the relationship between past value added and principal ratings and future teacher value-added using empirical Bayes estimates of teacher value added. The empirical Bayes method "shrinks" teacher effect estimates toward the population mean, with the degree of shrinkage proportional to the standard error of the teacher effect estimate (see Morris (1983)). Jacob and Lefgren (2008) argue that estimation error in the teacher effects will produce attenuation bias when teacher effects are used as an explanatory variable in a regression context. However, we obtain results similar to those reported in Table 12 when we use Empirical Bayes estimates in place of the non-shrunken teacher fixed effects.

late spring of 2006, based on whether they had student achievement data for the most recent year available at that time (the 2004/05 school year). A teacher with achievement data for 2004/05 who subsequently left the school (or who switched to a non-tested grade and subject) would therefore be excluded from the 2005/06 value-added sample. Rather than toss out a large proportion of the sample for all years, we report one-year value-added estimates for 2005/06 using the reduced sample. To distinguish between sample-size effects and differences due to the number of years of student achievement used to estimate teacher value-added, we also report results using single-year value-added estimates for 2004/05, which utilize the full sample of teachers.

It is not obvious a priori which of the measures should be the best predictors of future value-added, even when past value-added includes six years of information. On the one hand, we would expect value-added based on fewer years of prior information to be less precise, both because it is based on fewer student test scores and because it is subject to non-persistent changes in the student and teacher performance (e.g. a particularly disruptive student during the school year, student illnesses on exam day or temporary teacher health and family issues). On the other hand, actual (persistent) teacher performance could change over time, in which case teacher value-added from six years ago may not be very informative about the future, e.g., if a teacher received tenure four years ago and reduced effort thereafter then the two years of value-added information prior to the tenure decision will be misleading.

Our results support the first hypothesis, that random error is the key factor driving year-to-year variation in performance measures. Using all available information, past value-added outperforms principal ratings, explaining over five times as much of the variation in future value-added among math teachers and nearly 30 times as much of the variation in future value-added among reading teachers.[26] The edge in explanatory power (as measured by R-squared) holds up when only three, two or even a single year of data is used to compute past value added though the differential generally falls as value added is computed from fewer years of data. Thus if only a single measure is employed, past value added is superior to principal evaluations in predicting future teacher value added.

When prior value-added and principal ratings are combined to predict future teacher performance, the contribution of principal ratings to the predictive power of the model also depends on the precision of the past value-added measure. When past value-added is based on all six years of achievement gain data before Summer 2006, principal ratings add virtually nothing to the predictive power of past value-added in math or reading. The same is true when three or two years of student achievement data are used to compute prior value added. The results are mixed when past-added is based on a single year of data. If data from 2004/05 (and the constant sample of teachers) are used, combining prior value-added with principal evaluations increases the proportion of variation in future value added that is explained from 6.6 percent to 9.1 percent in math and from 2.0 percent to 2.3 percent in reading, though it is not possible to reject the null that principal ratings are uncorrelated with future value added (conditional on past value added). When data from 2005/06 alone (and the associated smaller sample) are used to construct prior value-added, principal ratings (which are typically based on multiple years of observation) do have a statistically significant correlation with future value added, conditional on past value added, in math. However, the improvement in explanatory power (measured by R-squared) is modest, 18 percent to 22 percent. Similarly, adding principal evaluations to one-

[26] Similar results are obtained when we exclude achievement data from the 05/06 school year from the value-added calculation and use only four years of test-score data (as in Jacob and Lefren (2008)).

year past value-added boosts the ability to predict future value added slightly, from 11 percent to 12 percent.

V. Summary and Conclusions

Consistent with prior research, we find that estimates of teachers' contributions to student achievement or "value-added" are at best weakly correlated with readily observable teacher characteristics like attainment of advanced degrees, indicating that other factors may be relatively more important in determining teacher productivity. Prior economic research suggests that non-cognitive factors may be particularly important, and often overlooked, determinants of productivity in occupations like teaching. We find that teacher value-added is correlated with traditional human capital measures like teacher intelligence, subject knowledge and teaching skills in math, while personality traits like motivation and enthusiasm are associated with high productivity among reading teachers. It may be that the non-cognitive teacher skills and traits that we find to be associated with teacher productivity in generating student achievement also contribute to non-academic outcomes as well. For example, Chetty et al. (2011) find that kindergarten class quality (including teacher quality) has significant effects on later non-cognitive outcomes, which in turn are associated with higher earnings, even though the impact of kindergarten class quality on student test scores fades out by middle school. Similarly, Chetty, Friedman, and Rockoff (2011) find that the influence of high-value-added elementary teachers is strongly correlated with both earnings in adulthood and non-pecuniary choices later in life, like the probability of teenage child-bearing.

The fact that non-cognitive factors are sometimes related to teacher value-added suggests that direct observation of potential teachers during the hiring process and subsequent observation of teacher behavior in the classroom would better identify effective teachers than the traditional system of assessing teachers on the basis of credentials alone. Thus there are potential gains from giving school principals a greater role in evaluating teachers. In addition to their ability to capture non-cognitive skills, evaluations by principals have a potential cost advantage over other alternatives, like classroom observation by external evaluators (as in the MET project). Principals collect most of their information in the natural course of the job (e.g., informal conversations with parents, students, and other teachers), which makes the marginal cost low.

The relative importance of intelligence, subject knowledge and teaching skills in determining math teacher productivity has important implications for recruiting and preparing future teachers as well. Because of the apparent role of intelligence, this would seem to suggest that policies designed to reduce entry barriers and encourage the “brightest” into the teaching profession could boost student achievement. However, this is tempered by the fact that subject matter knowledge and teaching skills seem to matter as well. Sheer intelligence may not be enough; “good” teachers likely need to have adequate training in subject matter content and essential teaching techniques.

While principal evaluations of teachers incorporate traits associated with value-added, like teacher knowledge, skill and intelligence, they also include assessments of a teacher’s interpersonal relationships with parents, other teachers and the principal, as well as a caring attitude toward students. This divergence in the factors associated with teacher value-added and those which are related to principal evaluation suggest that principal evaluations of teachers may also be useful if educational objectives beyond student achievement are valued.

The ability of principals to distinguish differences in teacher productivity does appear to vary across principals and across the spectrum of teacher quality. Contrary to the prior work of Jacob and Lefgren, we find that principals are much better at identifying the least productive teachers than they are at determining those who are best at raising student test scores. Thus principal evaluations would appear to be more valuable in retention decisions than in allotting rewards under a performance-pay system. While principals vary in their ability to identify differences in teacher productivity, we could not identify consistent relationships between principals evaluation ability and either their experience as an administrator or their tenure in a school. Future research in this area, with larger samples, could explore the ways in which principal characteristics as well as organizational forms and hierarchies influence principals' ability to identify effective teachers.

Our analysis of the predictive power of principal ratings and past value-added also informs the current policy debate over the use of test scores and subjective evaluations to evaluate current teachers. When value-added measures are constructed from multiple years of test score data, past value-added does a much better job at predicting future value-added than do principal evaluations. However, if one only uses a single year of information to estimate prior teacher value-added, principal evaluations add some information, though the gains are modest. Thus subjective measures are likely to be of greatest value when making retention and tenure decisions, especially for early-career teachers, for whom there may be only a year or two of student test-score information.

While our analysis is informative regarding the various ways that teachers could be assessed, it is important to be cautious in drawing broad policy conclusions from these results. For example, while we have shown that prior value-added is the best predictor of future value-added, future value-added is not necessarily an accurate indicator of overall future teacher productivity.[27] Value-added is a noisy measure of a teacher's impact on current student achievement

[27] The same critique applies to the MET project, which investigates the relationship between value added and classroom observational assessments from trained raters.

and may not capture other valuable contributions a teacher makes to a student's long-run success.[28] Also, the fact that principals' assessments are positively related to future value-added, and sometimes add information beyond prior value added, does not mean that evaluating teachers based on principals' assessments would necessarily be a wise policy for high-stakes personnel decisions. The assessments that principals offered in our study involved no financial or employment implications for teachers and principals' stated judgments could well differ in a high-stakes context. Also, even if principals would give the same assessments in high-stakes settings, doing so could influence the working relationships between principals and teachers in unproductive ways.

While caution is warranted, the practical reality is that many school systems around the country are already making radical changes to the way in which teachers are evaluated and compensated. Our results suggest principal evaluations can be a useful component of these new teacher assessment systems. First, employing principal evaluations is likely to be superior to the traditional credential and seniority based system of compensating teachers. Second, in systems where "value-added" metrics are used, including principal evaluations will be most informative for early-career teachers (where value-added information is less reliable). Third, because principals appear to be better at identifying the least productive teachers rather than the top performers, principal evaluations are more likely to be a reliable factor in termination decisions rather than in performance-pay systems. Finally, because principal evaluations take into account a broader set of teacher traits than those which directly affect student test scores, evaluations of teachers by principals are likely to be a useful component of teacher assessment when outcomes beyond student achievement are valued.

[28] For evidence on the relationship between value-added and long-run student outcomes, see Chetty, Friedman and Rockoff (2011).

References

Armor, David, Patricia Conry-Oseguera, Millicent Cox, Nicelma King, Lorraine McDonnell, Anthony Pascal, Edward Pauly and Gail Zellman (1976). *Analysis of the School Preferred Reading Program in Selected Los Angeles Minority Schools*. Report #R-2007-LAUSD. Santa Monica, CA: RAND Corporation.

Aaronson, Daniel, Lisa Barrow and William Sander (2007). "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics* 25(1): 95-135.

Baker, E.L., Barton, P.E., Darling-Hammond, L., Haertel, E., Ladd, H.F., Linn, R.L., Ravitch, D., Rothstein, R., Shavelson, R.J., Shepard, L.A. (2010). *Problems with the Use of Student Test Scores to Evaluate Teachers*. Washington, DC: The Economic Policy Institute.

Bommer, William H., Jonathan L. Johnson, Gregory A. Rich, Philip M. Podsakoff, and Scott B. MacKenzie (1995). "On the Interchangeability of Objective and Subjective Measures of Employee Performance: a Meta-analysis." *Personnel Psychology* 48(3): 587-605.

Borghans, Lex, Angela Lee Duckworth, James J. Heckman, and Bas ter Weel (2008). "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43(4): 972–1059.

Borghans, Lex, Bas ter Weel, and Bruce A. Weinberg (2008). "Interpersonal Styles and Labor Market Outcomes." *Journal of Human Resources* 43(4): 815–858.

Boyd, Donald, Pamela Grossman, Hamilton Lankford, Susanna Loeb and James Wyckoff. 2006. "How Changes in Entry Requirements Alter the Teacher Workforce and Affect Student Achievement." *Education Finance and Policy* 1(2):176-216.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach and Danny Yagan (2011). "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126: 1593-1660.

Chetty, Raj, John N. Friedman and Jonah E. Rockoff (2011). "The Long-term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood." Working paper no. 17699. Cambridge, MA: National Bureau of Economic Research.

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

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor (2007). "Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects." *Economics of Education Review* 26: 673-682.

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor (2010). "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects." *Journal of Human Resources* 45(3): 655-681.

Cunha, Flavio, James Heckman, Lance Lochner and Dimitry Masterov (2006). "Interpreting the Evidence on Life Cycle Skill Formation," In Handbook of the Economics of Education. ed, Eric A. Hanushek and Frank Welch. 697-812. Amsterdam: North-Holland.

Figlio, David N. and Lawrence W. Kenny (2007). "Individual Teacher Incentives and Student Performance." *Journal of Public Economics* 91: 901-914.

Gallagher, H. Alix (2004). Vaughan Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement. *Peabody Journal of Education*, 79(4): 79-107.

Goldhaber, Dan (2007). “Everyone’s Doing It, but What Does Teacher Testing Tell Us about Teacher Effectiveness?” *Journal of Human Resources* 42(4): 765-794.

Gordon, Robert, Thomas J. Kane and Douglas O. Staiger (2006). “Identifying Effective Teachers Using Performance on the Job.” Discussion Paper 2006-01. Washington, DC: The Brookings Institution.

Guarino, Cassandra M., Mark Reckase and Jeffrey Wooldridge (2011). “Can Value-Added Measures of Teacher Performance be Trusted?” Unpublished manuscript.

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

Hanushek, Eric A. (1997). “Assessing the Effects of School Resources on Student performance: an Update.” *Educational Evaluation and Policy Analysis* 19(2): 141-164.

Hanushek, Eric A., John F. Kain, Daniel M. O’Brien, and Steven G. Rivkin (2005). “The Market for Teacher Quality.” NBER Working Paper #11154.

Harcourt Assessment (2002). “SAT-10 to SAT-9 Scaled Score to Scaled Score Conversion Tables.”

Harris, Douglas N. (2011). *Value-Added Measures in Education*. Cambridge, MA: Harvard Education Press.

Harris, Douglas N., Carolyn Herrington and A. Albee (2007). “The Future of Vouchers: Lessons from the Adoption, Design, and Court Challenges of Florida’s Three Voucher Programs.” *Educational Policy*.

Harris, Douglas N., Stacey Rutledge, William Ingle, and Cynthia Thompson (2010). "Mix and match: What principals really look for when hiring teachers." *Education Finance and Policy* 5(2): 228-246.

Harris, Douglas N., William Ingle, and Stacey Rutledge (2012). "How Teacher Evaluation methods Matter for Accountability." Unpublished working paper.

Harris, Douglas N., Tim R. Sass and Anastasia Semykina (2011). "Value-Added Models and the Measurement of Teacher Productivity." Unpublished. Tallahassee, FL: Florida State University.

Harris, Douglas N. and Tim R. Sass (2011). "Teacher Training, Teacher Quality and Student Achievement." *Journal of Public Economics* 95: 798-812.

Heckman, James J., Jora Stixrud, Sergio Urzua (2006) The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3), 411-482.

Heneman, Robert L. (1986). "The Relationship Between Supervisory Ratings and Results-Oriented Measures Performance: a Meta-analysis." *Personnel Psychology* 39: 811-826.

Jacob, Brian A., and Lars Lefgren (2005). "Principals as Agents: Subjective Performance Measurement in Education." Working Paper #11463. Cambridge, MA: National Bureau of Economic Research.

Jacob, Brian A. and Lars Lefgren (2008). "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education." *Journal of Labor Economics*, 26(1): 101-136.

Jepsen, Christopher (2005). "Teacher Characteristics and Student Achievement: Evidence from Teacher Surveys." *Journal of Urban Economics* 57(2):302-19.

Kane, Thomas J., Douglas O. Staiger, Daniel McCaffrey, Steve Cantrell, Jeff Archer, Sarah Bugayar, Kerri Kerr, Todd Kawakita and David Parker. *Gathering Feedbck for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains*. Seattle, WA: Bill and Melinda Gates Foundation.

Kane, Thomas J., Jonah E. Rockoff and Douglas O. Staiger (2008). "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." *Economics of Education Review* 27: 615-631.

Kimball, S. M., & Milanowski, A. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education,* 79(4), 54-78.

Koedel, Cory. 2009. "An Empirical Analysis of Teacher Spillover Effects in Secondary School." *Economics of Education Review* 28(6): 682-692.

Koedel, Cory, and Julian R. Betts. (2007). "Re-Examining the Role of Teacher Quality In the Educational Production Function." Unpublished working paper, University of Missouri.

Koedel, Cory and Julian R. Betts (2011). "Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique." *Education Finance and Policy* 6(1): 18-42.

Lewis, Jeffrey B. and Drew A. Linzer (2005). "Estimating Regression Models in Which the Dependent Variable Is Based on Estimates." *Political Analysis*, 13(4): 345–364. .

McCaffrey, Daniel F., Tim R. Sass, J.R. Lockwood and Kata Mihaly (2009). "The Inter-temporal Variability of Teacher Effect Estimates. *Education Finance and Policy*, 4(4), 572-606.

Medley, Donald M. and Homer Coker (1987). "The Accuracy of Principals' Judgments of Teacher Performance." *Journal of Educational Research* 80(4): 242-247.

Mihaly, Kata, Daniel F. McCaffrey, J.R. Lockwood and Tim R. Sass (2010). "Centering and Reference Groups for Estimates of Fixed Effects: Modifications to felsdvreg." *Stata Journal*, 10(1), 82-103.

Milanowski, A. (2004). "The relationship between teacher performance evaluation scores and student assessment: Evidence from Cincinnati." *Peabody Journal of Education*, 79(4): 33-53.

Morris, Carl N. (1983). "Practical Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association* 78(381): 47-55.

Moulton, Brent R. (1990). "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *Review of Economics and Statistics* 72(2):334–38.

Murnane, Richard J. (1975). *The Impact of School Resources on the Learning of Inner City Children*. Cambridge, MA: Ballinger.

National Center for Education Statistics. (2006). *Characteristics of Schools, Districts, Teachers, Principals, and School Libraries in the United States: 2003-04 Schools and Staffing Survey*. Washington, DC.

Podgursky, Michael J. and Matthew G. Springer (2007). "Teacher Performance Pay: A Review." *Journal of Policy Analysis and Management*, 26(4): 909-949.

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

Rockoff, Jonah E. (2004). "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94(2): 247-52.

Rockoff, Jonah E., Douglas O. Staiger, Thomas J. Kane and Eric S. Taylor (forthcoming). "Information and Employee Evaluation: Evidence from a

Randomized Intervention in Public Schools." *American Economic Review*.

Rothstein, Jesse (2010). "Teacher Quality in Educational Production: Tracking, Decay and Student Achievement." *Quarterly Journal of Economics* 125(1): 175-214.

Todd, Petra E. and Kenneth I. Wolpin (2003). "On the Specification and Estimation of the Production Function for Cognitive Achievement." *Economic Journal* 113(485):F3-F33.

Varma, Arup and Linda K. Stroh (2001). "The Impact of Same-sex LMX Dyads on Performance Evaluations." *Human Resource Management* 40(4): 309-320.

Xu, Zeyu, Jane Hannaway and Colin Taylor. 2011. "Making a Difference? The Effects of Teach for America in High School." *Journal of Policy Analysis and Management* 30(3):447-469.

Table 1
Sample Student and Teacher Characteristics

	Math Sample		Reading Sample	
	No. of Obs.	Mean	No. of Obs.	Mean
Students				
Black	31645	0.367	30794	0.360
Hispanic	31645	0.025	30794	0.024
Free/Reduced Price Lunch	31645	0.335	30794	0.329
Achievement Gain	31645	20.729	30794	18.581
Teachers				
Male	1023	0.115	1024	0.079
White	1023	0.695	1024	0.724
Hold Advanced Degree	1004	0.332	1008	0.350
Fully Certified	1015	0.950	1019	0.955
Taught Primarily Elementary School	1023	0.727	1024	0.729
Taught Primarily Middle School	1023	0.149	1024	0.141
Taught Primarily High School	1023	0.124	1024	0.130
Principal’s Overall Rating	237	7.084	231	7.134
Rating of Ability to Raise Test Scores	210	7.200	201	7.184
Rating on “Caring”	237	7.384	231	7.463
Rating on “Enthusiastic”	237	7.249	231	7.372
Rating on “Motivated”	237	7.414	231	7.481
Rating on “Strong Teaching Skills”	237	7.544	231	7.636
Rating on “Knows Subject”	237	7.848	231	7.918
Rating on “Communication Skills”	237	7.612	231	7.758
Rating on “Intelligence”	237	7.911	231	7.970
Rating on “Positive Relationship with Parents”	236	7.483	230	7.600
Rating on “Positive Relationship with Students”	236	7.636	230	7.739

Note: Includes only students and teachers for which a fixed effect could be computed for the teacher.

Table 2
Factor Loadings of Normalized Principal Ratings

Teacher Characteristic Rated by Principal	Interpersonal Skills	Motivation/ Enthusiasm	Works Well With Others	Knowledge/ Teaching Skills/ Intelligence
		Math		
Intelligent	-0.0481	0.0839	0.0606	**0.7067**
Works Well With Grade Team/Dept.	-0.0046	-0.0887	**0.9711**	0.0399
Works Well With Me (Principal)	0.1743	0.0835	**0.7415**	-0.0814
Positive Relationship With Parents	**0.7231**	0.0781	0.0768	0.0742
Positive Relationship With Students	**0.9408**	0.0103	-0.0131	0.0636
Caring	**0.5591**	0.1372	0.2422	-0.0185
Enthusiastic	0.1086	**0.9721**	-0.0707	-0.0035
Motivated	0.0398	**0.5224**	0.2802	0.1624
Strong Teaching Skills	0.1512	0.0258	-0.0462	**0.8471**
Knows Subject	-0.0088	-0.0551	-0.0036	**0.9831**
Communication Skills	0.1040	0.1705	0.2734	0.3191
		Reading		
Intelligent	-0.0138	0.0094	0.0445	**0.7064**
Works Well With Grade Team/Dept.	0.0179	-0.0581	**0.8646**	0.0704
Works Well With Me (Principal)	0.1507	0.0409	**0.8251**	-0.0558
Positive Relationship With Parents	**0.7559**	0.0511	0.0637	0.0741
Positive Relationship With Students	**0.9195**	0.0258	0.0181	0.0287
Caring	**0.5970**	0.0989	0.2610	-0.0385
Enthusiastic	0.0728	**0.9942**	-0.0476	-0.0225
Motivated	0.0728	**0.5289**	0.1894	0.2529
Strong Teaching Skills	0.2269	0.0127	-0.0854	**0.8175**
Knows Subject	-0.0814	-0.0201	0.0333	**0.9840**
Communication Skills	0.1484	0.2225	0.1855	0.3214

Note: Principal ratings are normalized within principal to have mean zero and variance of one. Factor analysis uses maximum likelihood method. Factor loadings based on promax rotation.

Table 3
Pairwise Correlation of Estimated Teacher Fixed With Teacher Characteristic Factors

	Estimated Teacher FE	Inter-personal Skills	Moti-vation/ Enthus-iasm	Works Well With Others	Knowledge/ Teaching Skills/ Intelligence
			Math		
Estimated Teacher FE	1.000				
Interpersonal Skills	0.174** [0.194]	1.000			
Motivation/Enthusiasm	0.202** [0.225]	0.734**	1.000		
Works Well With Others	0.189** [0.211]	0.756**	0.732**	1.000	
Knowledge/Teaching Skills/ Intelligence	0.309** [0.344]	0.612**	0.682**	0.644**	1.000
			Reading		
Estimated Teacher FE	1.000				
Interpersonal Skills	0.166** [0.223]	1.000			
Motivation/Enthusiasm	0.336** [0.450]	0.631**	1.000		
Works Well With Others	0.222** [0.298]	0.716**	0.683**	1.000	
Knowledge/Teaching Skills Intelligence	0.153** [0.205]	0.762**	0.650**	0.676**	1.000

Note: **indicates significance at the .05 level. Correlations adjusted for estimation error in estimated teacher fixed effects are in brackets.

Table 4
FGLS Estimates of the Relationship Between
Teacher Fixed Effects and Teacher Characteristic Factors
(Grades 2 – 10, 1999/2000 – 2005/06)

	Math		Reading	
	[1]	[2]	[1]	[2]
Interpersonal Skill	-0.004 (0.023)		-0.010 (0.015)	
Knowledge/Teaching Skills/ Intelligence	0.058*** (0.019)		-0.009 (0.014)	
Motivation/Enthusiasm	0.008 (0.022)		0.042*** (0.013)	
Works Well With Others	-0.001 (0.023)		0.014 (0.014)	
Interpersonal Skill × Elementary		-0.009 (0.025)		-0.006 (0.017)
Interpersonal Skill × Middle/High		0.018 (0.058)		-0.023 (0.037)
Knowledge/Teaching Skills/ Intelligence × Elementary		0.057*** (0.022)		-0.013 (0.016)
Knowledge/Teaching Skills/ Intelligence × Middle/High		0.054 (0.044)		0.007 (0.035)
Motivation/Enthusiasm × Elementary		0.007 (0.025)		0.043*** (0.014)
Motivation/Enthusiasm × Middle/High		0.027 (0.063)		0.044 (0.037)
Works Well With Others × Elementary		0.008 (0.026)		0.017 (0.016)
Works Well With Others × Middle/High		-0.046 (0.059)		0.003 (0.026)
R-squared	0.137	0.140	0.145	0.148
No. of Observations	207	207	203	203

Note: Standard errors appear in parentheses. * indicates statistical significance at .10 level, **indicates significance at the .05 level and *** indicates significance at the .01 level in a two-tailed test. All models include controls for teacher experience, attainment of an advanced degree and a constant term.

Table 5
Pairwise Correlation of Principal Ratings of Teachers With Teacher Characteristic Factors

	Overall Rating	Ability to Raise Test Scores	Inter-personal Skills	Moti-vation/ Enthus-iasm	Works Well With Others	Knowledge/ Teaching Skills/ Intelligence
			Math			
Overall Rating	1.000					
Ability to Raise Test Scores	0.733**	1.000				
Interpersonal Skills	0.703**	0.550**	1.000			
Motivation/Enthusiasm	0.738**	0.596**	0.734**	1.000		
Works Well With Others	0.762**	0.598**	0.756**	0.732**	1.000	
Knowledge/Teaching Skills/ Intelligence	0.881**	0.752**	0.612**	0.682**	0.644**	1.000
			Reading			
Overall Rating	1.000					
Ability to Raise Test Scores	0.741**	1.000				
Interpersonal Skills	0.709**	0.626**	1.000			
Motivation/Enthusiasm	0.856**	0.702**	0.631**	1.000		
Works Well With Others	0.697**	0.569**	0.716**	0.684**	1.000	
Knowledge/Teaching Skills Intelligence	0.723**	0.589**	0.763**	0.650**	0.676**	1.000

Note: **indicates significance at the .05 level. Correlations adjusted for estimation error in estimated teacher fixed effects are in brackets.

Table 6
Ordinary Least Squares Estimates of the Relationship Between
Principal Overall Ratings of Teachers and Teacher Characteristic Factors
(Grades 2 – 10, 1999/2000 – 2005/06)

	Math		Reading	
	[1]	[2]	[1]	[2]
Interpersonal Skill	0.096** (0.047)		0.187*** (0.056)	
Knowledge/Teaching Skills/ Intelligence	0.608*** (0.040)		0.156*** (0.054)	
Motivation/Enthusiasm	0.054 (0.046)		0.601*** (0.048)	
Works Well With Others	0.233*** (0.047)		0.024 (0.052)	
Interpersonal Skill × Elementary		0.108** (0.052)		0.130** (0.062)
Interpersonal Skill × Middle/High		0.051 (0.124)		0.505*** (0.137)
Knowledge/Teaching Skills/ Intelligence × Elementary		0.615*** (0.045)		0.190*** (0.058)
Knowledge/Teaching Skills/ Intelligence × Middle/High		0.599*** (0.096)		-0.031 (0.140)
Motivation/Enthusiasm × Elementary		0.043 (0.050)		0.613*** (0.050)
Motivation/Enthusiasm × Middle/High		0.176 (0.135)		0.440*** (0.145)
Works Well With Others × Elementary		0.248*** (0.052)		0.056 (0.058)
Works Well With Others × Middle/High		0.112 (0.126)		-0.048 (0.108)
R-squared	0.852	0.854	0.805	0.814
No. of Observations	207	207	203	203

Note: Standard errors appear in parentheses. * indicates statistical significance at .10 level, **indicates significance at the .05 level and *** indicates significance at the .01 level in a two-tailed test. All models include controls for teacher experience, attainment of an advanced degree and a constant term.

Table 7
Pairwise Correlation of Estimated Teacher Fixed Effects,
Principal's Rating of Teachers and Average Student Developmental Scale Score Gains
(Teachers with Students who took FCAT-SSS exam in 2003/04 and 2004/05 or 2004/05 and 2005/06)

	Estimated Teacher FE	Overall Rating	Ability to Raise Test Scores	Average Dev. Scale Score Gain (2004/05)	Average Dev. Scale Score Gain (2005/06)
Math					
Estimated Teacher FE	1.000				
Overall Rating	0.204** [0.232]	1.000			
Ability to Raise Test Scores	0.273** [0.310]	0.674**	1.000		
Average DSS Gain (2004/05)	0.189** [0.214]	0.189**	0.108	1.000	
Average DSS Gain (2005/06)	0.268** [0.304]	0.055	0.121	0.370**	1.000
Reading					
Estimated Teacher FE	1.000				
Overall Rating	0.190** [0.244]	1.000			
Ability to Raise Test Scores	0.142 [0.182]	0.676**	1.000		
Average DSS Gain (2004/05)	-0.401 [-0.052]	0.146	0.063	1.000	
Average DSS Gain (2005/06)	0.128 [0.164]	0.088	0.069	0.495**	1.000

Note: **indicates significance at the .05 level. Correlations adjusted for estimation error in estimated teacher fixed effects are in brackets.

Table 8
FGLS Estimates of the Determinants of Teacher Fixed Effects
(Grades 2 – 10, 1999/2000 – 2005/06)

	Math			Reading		
	[1]	[2]	[3]	[1]	[2]	[3]
Overall Rating		0.059*** (0.013)			0.038*** (0.008)	
Ability to Raise Test Scores			0.066*** (0.034)			0.031*** (0.009)
1-2 Years of Experience	0.048 (0.176)	0.105 (0.169)	0.076 (0.181)	0.027 (0.108)	0.025 (0.104)	0.054 (0.146)
3-5 Years of Experience	0.089 (0.128)	0.103 (0.123)	0.095 (0.131)	0.064 (0.082)	0.051 (0.078)	0.070 (0.111)
6-12 Years of Experience	0.076 (0.124)	0.115 (0.119)	0.086 (0.127)	0.052 (0.080)	0.057 (0.077)	0.079 (0.110)
13-20 Years of Experience	0.038 (0.124)	0.082 (0.119)	0.076 (0.127)	0.041 (0.080)	0.046 (0.077)	0.060 (0.111)
21-27 Years of Experience	0.134 (0.125)	0.158 (0.120)	0.136 (0.126)	0.085 (0.080)	0.072 (0.077)	0.106 (0.111)
28+ Years of Experience	0.092 (0.127)	0.134 (0.122)	0.093 (0.130)	0.078 (0.081)	0.077 (0.078)	0.093 (0.112)
Advanced Degree	-0.022 (0.026)	-0.025 (0.025)	-0.013 (0.027)	-0.001 (0.017)	-0.003 (0.016)	-0.010 (0.018)
R-squared	0.026	0.110	0.137	0.025	0.114	0.091
No. of Observations	237	237	202	231	231	201

Note: Standard errors appear in parentheses. * indicates statistical significance at .10 level, **indicates significance at the .05 level and *** indicates significance at the .01 level in a two-tailed test. All models include a constant term.

Table 9
FGLS Estimates of the Determinants of Teacher Fixed Effects,
Allowing for Differential Effects by Grade Group
(Grades 2 – 10, 1999/2000 – 2005/06)

	Math			Reading		
	[1]	[2]	[3]	[1]	[2]	[3]
Overall Rating × Elementary		0.060***			0.038***	
		(0.015)			(0.009)	
Overall Rating × Middle/High		0.053**			0.040**	
		(0.025)			(0.016)	
Ability to Raise Test Scores × Elementary			0.070***			0.044***
			(0.016)			(0.010)
Ability to Raise Test Scores × Middle/High			0.055**			0.000
			(0.026)			(0.016)
1-2 Years of Experience	0.048	0.104	0.078	0.027	0.024	0.105
	(0.176)	(0.170)	(0.181)	(0.108)	(0.104)	(0.146)
3-5 Years of Experience	0.089	0.101	0.094	0.064	0.051	0.096
	(0.128)	(0.124)	(0.132)	(0.082)	(0.079)	(0.110)
6-12 Years of Experience	0.076	0.115	0.085	0.052	0.056	0.114
	(0.124)	(0.120)	(0.127)	(0.080)	(0.077)	(0.110)
13-20 Years of Experience	0.038	0.081	0.074	0.041	0.045	0.091
	(0.124)	(0.120)	(0.127)	(0.080)	(0.077)	(0.110)
21-27 Years of Experience	0.134	0.157	0.134	0.085	0.071	0.135
	(0.125)	(0.120)	(0.127)	(0.080)	(0.077)	(0.110)
28+ Years of Experience	0.092	0.133	0.094	0.078	0.076	0.130
	(0.127)	(0.123)	(0.130)	(0.081)	(0.078)	(0.112)
Advanced Degree	-0.023	-0.023	-0.013	-0.001	-0.004	-0.005
	(0.026)	(0.025)	(0.027)	(0.017)	(0.016)	(0.017)
R-squared	0.026	0.110	0.138	0.025	0.114	0.116
No. of Observations	237	237	202	231	231	201

Note: Standard errors appear in parentheses. * indicates statistical significance at .10 level, **indicates significance at the .05 level and *** indicates significance at the .01 level in a two-tailed test. All models include a constant term.

Table 10
FGLS Estimates of the Determinants of the
Correlation Between Teacher Fixed Effects and Principal Evaluations

	Math			Reading		
	[1]	[2]	[3]	[1]	[2]	[3]
Correlation Between Teacher Effects and Overall Rating of Teacher						
Principal's Tenure at School	0.021		0.039*	-0.008		0.006
	(0.020)		(0.023)	(0.021)		(0.026)
Principal's Total Experience in Ed. Administration		-0.009	-0.023		-0.012	-0.014
		(0.014)	(0.015)		(0.013)	(0.016)
R-squared	0.042	0.019	0.124	0.008	0.043	0.046
No. of Observations	26	26	26	22	22	22
Correlation Between Teacher Effects and "Ability to Raise Test Scores"						
Principal's Tenure at School	-0.029		-0.045**	0.008		0.009
	(0.019)		(0.023)	(0.020)		(0.027)
Principal's Total Experience in Ed. Administration		0.001	0.024		0.003	-0.002
		(0.017)	(0.020)		(0.017)	(0.023)
R-squared	0.010	0.015	0.155	0.008	0.002	0.008
No. of Observations	24	24	24	22	22	22

Note: Standard errors appear in parentheses. * indicates statistical significance at .10 level, **indicates significance at the .05 level and *** indicates significance at the .01 level in a two-tailed test. All models include a constant term. Data on principal placements are only available from 1995 forward, so tenure at school is truncated at 10 years.

Table 11
Rankings of Teacher Fixed Effects by Principal Ratings of Teachers

Math			
	Principal's "Overall Rating of Teacher" Percentile Ranking		
Teacher Fixed Effects Percentile Ranking	Bottom 30%	Middle 40%	Top 30%
Bottom 30%	**65**	24	11
Middle 40%	51	**32**	18
Top 30%	32	52	**16**
Reading			
	Principal's "Overall Rating of Teacher" Percentile Ranking		
Teacher Fixed Effects Percentile Ranking	Bottom 30%	Middle 40%	Top 30%
Bottom 30%	**54**	29	17
Middle 40%	41	**40**	19
Top 30%	34	45	**21**

Note: cell entries represent row percentages. Only schools with 5 or more rated teachers are included.

Table 12
FLGS Estimates of the Determinants of Teacher Effects in 2006/07 – 2007/08
(Only Teachers Teaching in Same School in Which They Were Previously Rated by Principal)

	Math			Reading		
	[1]	[2]	[3]	[1]	[2]	[3]
Prior Value-Added Based on Up to Six Years of Teacher Performance						
Prior Value-Added	0.562***		0.532***	0.724***		0.756***
(from 00/01-05/06)	(0.077)		(0.079)	(0.136)		(0.144)
Principal's Overall Rating		0.052***	0.025		0.019	-0.014
(Summer 2006)		(0.019)	(0.017)		(0.021)	(0.021)
R-squared	0.255	0.048	0.265	0.172	0.006	0.175
No. of Observations	158	158	158	138	138	138
Prior Value-Added Based on Up to Three Years of Teacher Performance						
Prior Value-Added	0.527***		0.507***	0.371***		0.377***
(from 03/04-05/06)	(0.067)		(0.070)	(0.081)		(0.085)
Principal's Overall Rating		0.052***	0.015		0.019	-0.005
(Summer 2006)		(0.019)	(0.017)		(0.021)	(0.021)
R-squared	0.286	0.048	0.290	0.132	0.006	0.133
No. of Observations	158	158	158	138	138	138
Prior Value-Added Based on Up to Two Years of Teacher Performance						
Prior Value-Added	0.537***		0.514***	0.527***		0.529***
(from 04/05-05/06)	(0.065)		(0.067)	(0.105)		(0.108)
Principal's Overall Rating		0.052***	0.022		0.019	-0.002
(Summer 2006)		(0.019)	(0.016)		(0.021)	(0.020)
R-squared	0.304	0.048	0.312	0.156	0.006	0.156
No. of Observations	158	158	158	138	138	138

Prior Value-Added Based on One Year of Teacher Performance (2004/05)						
Prior Value-Added	0.147***		0.123***	0.098*		0.092
(from 04/05)	(0.044)		(0.045)	(0.059)		(0.060)
Principal's Overall Rating		0.052***	0.039		0.019	0.015
(Summer 2006)		(0.019)	(0.019)		(0.021)	(0.021)
R-squared	0.066	0.048	0.091	0.020	0.006	0.023
No. of Observations	158	158	158	138	138	138
Prior Value-Added Based on One Year of Teacher Performance (2005/06)						
Prior Value-Added	0.343***		0.301***	0.282***		0.271***
(from 05/06)	(0.064)		(0.065)	(0.076)		(0.076)
Principal's Overall Rating		0.071***	0.048**		0.038	0.031
(Summer 2006)		(0.020)	(0.019)		(0.023)	(0.022)
R-squared	0.179	0.086	0.217	0.109	0.023	0.124
No. of Observations	132	132	132	114	114	114

Note: Standard errors appear in parentheses. * indicates statistical significance at .10 level, **indicates significance at the .05 level and *** indicates significance at the .01 level in a two-tailed test. All models include a constant term.