WHAT MAKES FOR A GOOD TEACHER AND WHO CAN TELL?*

by

Douglas N. Harris
Department of Educational Policy Studies
University of Wisconsin – Madison

Tim R. Sass
Department of Economics
Florida State University

Original Version: March 5, 2007
This version: September 6, 2010

Abstract

We examine several issues regarding the measurement of worker productivity, using a sample of school teachers: the association between subjectively determined teacher characteristics and teacher productivity, the factors affecting the correlation between principals' subjective assessments and measured productivity, the determinants of subjective ratings, and the ability of principals' assessments of teachers to predict future teacher performance. We find that principals' evaluations are better predictors of a teacher's contribution to student achievement (value-added) than are traditional approaches to teacher compensation that focus on experience and formal education. Further, teacher value-added is most closely associated with teachers' subject knowledge, teaching skill, intelligence and motivation, while subjective teacher ratings encompass a broader set of teacher personality traits. Principals' ratings of teachers are more closely aligned with value-added in elementary school and when the principal has more experience. Finally, while past teacher value-added predicts future teacher value-added, the principals' subjective ratings can provide additional information and substantially increase predictive power.

---

*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. John Gibson, William Ingle, Micah Sanders, Cynthia Thompson and Lisa Verdon provided valuable research assistance. 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 considerably heterogeneity in teacher productivity within and across schools (Rockoff (2004), Hanushek, et al. (2005), Rivkin, Hanushek and Kain (2005), Kane, Rockoff and Staiger (2006), Aaronson, Barrow and Sander (2007)). However, relatively little is known about what makes some teachers more productive than others in promoting student achievement.

Older cross-sectional studies of educational production functions found that the characteristics that form the basis for teacher compensation—graduate degrees and experience—are at best weak predictors of a teacher's contribution to student achievement (Hanushek, 1986, 1997). More recent estimates using panel data have determined that teacher productivity increases over the first few years of experience (Rockoff (2004), Clotfelter, Ladd and Vigdor (2006), Jepsen (2005), Rivkin, Hanushek and Kain (2005), Harris and Sass (2008), Aaronson Barrow and Sander (2007)), but little else in the way of observed teacher characteristics seems to consistently matter.[1] In short, while teachers significantly impact student achievement, the variation in teacher productivity is still largely unexplained by commonly measured 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 determine productivity. Recent work in labor economics suggests, for example, that personality

[1] Clotfelter, Ladd and Vigdor (2007, forthcoming), 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.

traits such as conscientiousness play an important role in determining labor productivity (Borghans, ter Weel, and Weinberg (2008), Cunha, Heckman, Lochner, and Masterov (2006); Heckman, Stixrud, and Urzua (2006)).

Unraveling the factors that determine 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 as signals to identify the most desired candidates in the hiring process. Consistent with this notion, 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. If, however, valuable teacher characteristics are malleable, determining which teacher characteristics have the greatest impact on student learning could also inform the design of pre-service and in-service teacher training programs.

Understanding the factors that affect teacher productivity would also inform current policy debates over how best to evaluate and compensate teachers. If it is not possible to directly measure the characteristics of teachers that determine their productivity, then ex-post evaluation of teachers based on their contributions to student achievement, or "value added," may be optimal (Gordon, Kane, and Staiger (2006)). Spurred on by the federal *Teacher Incentive Fund* and *Race to the Top* initiatives, many school districts are experimenting with pay-for-performance systems that tie compensation to teacher value-added (Figlio and Kenny (2007), Podgursky and Springer (2007)). Numerous districts are also placing greater emphasis on value-added outcomes when making decisions about tenure and promotion. 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. Alternatively, if there are teacher characteristics and behaviors that influence teacher productivity and are observable, but perhaps not objectively quantifiable, then reliance on supervisor evaluations and other more subjective assessments may be advantageous. The downside of subjective evaluations by principals is they may be affected by personal bias toward factors unrelated to productivity and some principals may simply be poor judges of teacher productivity.

To address the interrelated issues of the determinants of teacher productivity and how best to evaluate teacher performance, we analyze the relationship between principal evaluations of teachers and the contribution of teachers to student achievement or teacher "value added." Similar to other recent studies (Jacob and Lefgren (2008), Rockoff, et al. (2010)) we analyze the correlation between teacher value-added scores and principals' subjective assessments and how the relationship between the two measures varies with the characteristics of the principal.

We build on extant research in three ways, however. First, we go beyond general ratings of teacher ability and estimate the relationship between a variety of specific teacher personality traits and our multiple measures of teacher productivity. Second, since the goal of evaluation for decisions like tenure is to predict future teacher performance, we test how well prior value-added scores and prior principal evaluations predict *future* teacher value-added, rather than just contemporaneous student achievement. Finally, 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.

In the next section, we describe the small existing literature on principal evaluations of teachers and their relationship with value added. This is followed by a discussion of the data used for our new 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 role of personality traits is also related to the way in which overall worker productivity is measured and rewarded in the workplace—in particular, the balance of subjective supervisor ratings and more objective measures of output. 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. 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.

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.[2] They find that principals can generally identify teachers who contribute the most and the least to student achievement, but are less able to distinguish teachers in the middle of the productivity distribution.

Jacob and Lefgren also find that previous value-added is a better predictor of current student outcomes 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 principal ratings were also significant predictors of current test scores,

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

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 stable indicators of previous teacher productivity, 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 a working paper by Jacob and Lefgren (2005). Their list of items rated by principals includes both teacher characteristics/inputs (dedication/work ethic, organization, classroom management, providing a role model for students, positive relationships with teacher colleagues and administrators) and outputs (raising student achievement, student and parent satisfaction). They also apply factor analysis to these variables and create three broader variables: student satisfaction, achievement, and collegiality. However, the teacher's relationship with the school administration is the only teacher characteristic they consider as a possible predictor of value-added. (Their evidence suggests a positive and significant relationship between the two.)

Rockoff, et al. (2010) 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 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 after the principal received value-added information and there was a corresponding 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 obtain positive estimated correlations, similar in magnitude to those obtained by Jacob and Lefgren. The correlations tend to increase the more precise the value-added estimates and the longer a principal has known a teacher.

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)).[3] 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.

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

[3] 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.

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 diverse 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 diversity 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 percentage with at least a master's degree (sample district: 100 percent; national: 90.7 percent).[4] 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 is now coming to pass in other states as a result of *No Child Left Behind.* 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 and the lowest grade was a D (one school). It is reasonable to expect that accountability policies, such as the school

[4] 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.

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.

B. Student Achievement Data and Modeling

Throughout Florida there is annual testing in grades 3-10 for both math and reading. Two tests are administered, a criterion-referenced exam based on the state curriculum standards known as the FCAT-Sunshine State Standards exam, and a norm referenced test which is the Stanford Achievement Test. We employ the Stanford Achievement Test 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 Stanford Achievement Test in grades 1 and 2, allowing us compute achievement gains for students in grades 2-10. Achievement data on the Stanford Achievement Test are available for each of the school years 1999/00 through 2007/08.[5] Thus we are able to estimate the determinants of achievement gains for five years prior to the

[5] 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).

principal interviews, 2000/01-2005/06, and for two years after the interviews, 2006/07-2007/08. 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 of the following form:

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

The vector $\mathbf{X}_{it}$ includes time varying student characteristics such as student mobility. 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. There are three fixed effects in the model: a student fixed effect (γ_i), a teacher fixed effect (δ_k), and a school fixed effect, ϕ_m. 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 model is based on the cumulative achievement model of Todd and Wolpin (2003) and it is derived in detail in Harris and Sass (2006).

Recently, Rothstein (2010) has argued that such value-added models may produce biased estimates of teacher productivity due to the non-random assignment of students to teachers within schools. While our use of student fixed effects controls for sorting based on time-invariant characteristics, Rothstein argues that teacher effect estimates could still be biased if teacher assignments are determined by transient shocks to student achievement. 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 (2009). 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.[6]

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 and subjective evaluations by principals. Third, in regression analyses where value-added is the dependent variable we use a feasible generalized least squares estimation procedure which accounts for error in the dependent variable.[7] 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

[6] For math, the p-values on the test of zero future teacher "effects" were 1.00 for elementary school, 0.75 for middle school and 0.63 for high school. For reading the corresponding p-values were 1.00, 0.35 and 0.20.

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

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 measurent error adjustments.

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 teacher 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 prior to conducting 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.

The first question in the interview involved asking the principals to mark on a sheet of paper the principal's overall assessment of each teacher, using a 1-9 scale.[8] 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.[9] Having an 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)).

[8] 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."

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

The interview questions were designed so that principals would evaluate teachers relative to others in the school.[10] 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, in our analyses 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, raises FCAT writing achievement, positive relationship with parents, and positive relationship with students. 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 an educational outcome or to difficulty in identifying teachers who increase student test scores.

Finally, as part of the interview, we discovered that principals have access to a district-purchased software program, Snapshot™, that allows them to create various cross-tabulations of student achievement data, 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

[10] In contrast, in the Rockoff, et al. (2010) 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.

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.

IV. Results

In order to compute value-added scores for teachers we estimate equation (1) using data on test score gains for grades 2-10 over the period 2000/01 through 2005/06. In order 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.

A. The Association of Principal Evaluations with Teacher Value-Added Performance

Simple correlations between the estimated teacher fixed effects, principals' overall ratings of teachers, principals' ratings of a teacher's ability to raise test scores on the relevant achievement test, 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, using the methodology employed by Jacob and Lefgren.[11] Looking at the first column, there are positive relationships between teacher value-added and all of the teacher characteristic factors. The overall principal rating is positively associated with value-added though, as in previous studies, this relationship is relatively weak (adjusted correlation = 0.32 for math and

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

0.24 for reading). The correlation between value-added and the principal's impression of a teacher's ability to raise test scores (the subjective equivalent of value-added) is similarly low (adjusted correlation = 0.28 for math and 0.25 for reading).

The relatively high correlation of 0.7 between principals' overall rating and their ratings on ability of teachers to raise test scores suggests that principals weigh the ability of teachers to boost student test scores highly in their overall evaluation. These findings hold for both math and reading. It is also noteworthy that the teacher-characteristics factors are all positively correlated with one another, and often highly correlated. 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 the motivation/enthusiasm of the teacher.

Table 4 presents feasible generalized least squares (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, certification status) are included as explanatory variables. With the exception of one of the experience measures, none of the credential variables is a statistically significant

determinant of teacher value added scores.[12] None of the coefficients is significant in the first column for reading.[13]

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, educational attainment and certification that typically form the basis for teacher pay scales.

It is common to interpret the magnitude of coefficients in these types of models in terms of student-level standard deviations. For example, the coefficient on principals' overall ratings for math teachers in Table 4 is +3.136, which implies that a teacher who is rated one point higher on the 1-9 scale raises student math test scores by 3.1 scale score points per year more than the average teacher, which translates to 0.06 student test score standard deviations.[14] Put differently,

---

[12] In another study using statewide data from Florida (Harris and Sass (2008)), 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 large estimated coefficients here for full certification of reading teachers are likely picking up idiosyncratic features of the handful of reading teachers in the sample who are not fully certified during part of the sample period.

[13] The lack of significance may 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. Other studies using statewide Florida data do show positive coefficients on experience (Harris and Sass (2008)).

[14] This conversion is based on the standard deviation in the level of math achievement, 53.26. The standard deviation in the level of reading achievement is 50.58. Boyd, et al. (2008) argue for measuring teacher effects relative to the standard deviation of student gains. This would roughly double the effect sizes as the standard deviation of achievement gains are 23.20 for reading and 20.64 for math.

a one-standard deviation in increase in the principal's overall rating of a math teacher corresponds to a 0.1 standard deviation increase in the teacher's student's test scores. While this might be considered small by some standards, it represents only single-year changes, which could accumulate to relatively larger effects over time.

In addition to general correlations, we also consider the ability of principals to identify teachers at the extremes of the teacher effectiveness (value-added) distribution. In Table 5 we present cross-tabulations of the quintile rankings of teacher value added and principals' ratings of teachers on the "ability to raise test scores" metric. It appears that for both math and reading, principals are relatively better at identifying low value-added teachers, rather than top-performing teachers. Of those teachers who rank in the bottom quintile based on value-added in math, 41 percent are also ranked in the bottom quintile by their principal. In contrast, only nine percent of teachers in the top value-added quintile in math are also ranked in the top quintile by their principal. Similar differences appear for reading teachers.

B. Factors Affecting the Relationship Between Principal Ratings and Value-Added

In Table 6 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 the elementary level, the two principal ratings ("overall" and "ability to raise test scores") are positively and statistically significantly associated with the teacher fixed effect in both reading and in math. However, the effect of a one-point increase in the principal's rating scale on teacher value-added in reading is about half the size of the effect in 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 reading teachers.

For middle and high school teachers, there are no significant relationships.[15] This difference may reflect difficulties in aligning the content of exams with the teacher responsible for the relevant instruction in higher grade levels. In elementary schools, the matching of courses to the content of exams is relatively easy because students typically have only one teacher. In middle and high school, however, courses become more specialized and may cover lots of material other than what is tested on standardized exams. Also, instruction in other disciplines may affect student performance. For example, reading instruction and reading scores may be influenced by classes such as social studies, which involve reading but where developing reading is not the primary purpose.[16] Further, middle and high schools are generally much larger than elementary schools and thus the principal likely has less time to directly observe teacher performance.

As suggested by the divergent results between elementary and middle/high school, 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)

[15] While there are fewer middle and high school teachers than elementary teachers in the sample, the insignificant effects for middle/high school teachers are not due to the sample size. We restricted the sample so that the number of elementary teachers equaled the number of middle/high school teachers and obtained similar results.

[16] Koedel (2009) provides some evidence that social studies teachers influence reading test scores at the high school level.

and the principal's years of experience in educational administration (a proxy for general personnel evaluation skills). Results are presented in Table 7. We find some evidence to support the hypotheses that principal's ability to evaluate teachers is positively associated with both general administrative experience and with the extent of exposure to the teachers they rate. The evidence is weak, however, due to the relatively small sample size (25 principals in the "overall" ratings and 23 principals in the "ability to raise test scores" rankings).

C. Factors Determining Principals' Ratings of Teachers

We next turn to an analysis of the factors affecting a principal's overall rating of a teacher. Table 8 presents least squares estimates from regressing the principal's overall rating on the perceived ability to raise test scores in the relevant subject and the principal's overall rating of teachers. For both math and reading, ability to raise test scores is highly correlated with the overall rating. This is true for both all teachers as well as the subgroups of elementary and middle/high school teachers. However, there is more to the overall rating than ability to raise test scores; about 45 percent of the variation in overall ratings is due to other factors.

To determine what specific factors influence a principal's overall rating of a teacher we re-estimate the teacher rating model using the principal's rating of the four teacher characteristic factors. The results are presented in Table 9. In both subjects, knowledge/teaching skills/intelligence contributes the most to the principals' overall rating. "Works well with others" and interpersonal skill are also statistically significant but the point estimates are much smaller. There are some apparent differences by grade level, though none of these differences is statistically significant. Also, note that four factors explain roughly 80 percent of the variation in overall ratings, suggesting that the underlying 12 characteristics are important determinants of principals' overall ratings.

Very different patterns emerge when we switch the dependent variable to teacher value-added in Table 10. Column [1] suggests that knowledge/teaching skills/intelligence is positively and significantly associated with teacher value-added in reading and math. None of the other coefficients in column [1] are significant. Column [2] shows that the effect of knowledge/teaching/intelligence skills is entirely in the elementary grades. The overall explanatory power of the four factors is quite low, however.[17] For reading, the only factor which is even marginally significant is teacher motivation/enthusiasm. Once again, the effect is only significant for elementary school teachers.

Finally, we consider how principal ratings may be influenced by knowledge of student test scores. As mentioned above, 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 criterion-referenced exam known as the Florida Compreshensive Achivement Test, Sunshine State Standards or FCAT-SSS. This is a different exam than the norm-referenced FCAT-NRT we use to construct teacher value added throughout the paper. In Table 11 we show that principals' overall rankings of teachers are positively correlated with once-lagged average DSS 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 average student test-score gains in both math and reading. Thus while recent evidence produced by Rockoff et al in New York suggests that principals incorporate value-

[17] Some of the insignificant effects may be due to multicollinearity. As demonstrated in Table 5, 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.

added information in their assessment of teachers, in our Florida example we do not find strong evidence that principals are influenced by simple (not conditioned on student characteristics) average student test-score gains.

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

To this point, we have been using achievement data up through the 2005/06 school year to compute teacher value added and compared this value-added measure with various principal ratings of their teachers obtained in the Summer of 2006. Such contemporaneous estimates 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, which uses two estimated gains, 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. As shown in Table 12, we estimate the equation several ways, varying the amount of information used to estimate the past teacher value-added.[18]

The relative performance of principal ratings in predicting future value added varies directly with the amount of information used to compute prior value added. Using all available information, past value added outperforms principal ratings, explaining fourteen times as much of the variation in future value-added among math teachers and over twice as much of the variation among reading teachers. Similar results are obtained when we exclude the last year of achievement data from the value-added calculation (as in Jacob and Lefren (2008)). The edge in explanatory power holds up in math when only two prior years of data are used to compute past value added, but the difference is eliminated in reading. When past value added is based on a single year of data, principal ratings (which are typically based on multiple years of observation) outperform past value added in predicting the future value-added of math teachers. This reinforces the importance of using multiple years of data to estimate 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 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 and only slightly increase the proportion of the

[18] 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 standard teacher fixed effects.

explained variation in the future value-added of reading teachers from ten to 11 percent. As less data is used to construct prior value added estimates, the relative contribution of principal ratings grows. When a single year of student test-score gains is used to compute prior value added, principal ratings increase the proportion of variation in future value added explained by about two percentage points in math and nearly increase the proportion of variation explained in reading from 7 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 experience and attainment of advanced degrees, suggesting that other factors may be relatively more important in determining what makes a "good" teacher. 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. In contrast, principal evaluations of teachers appear to be based on a broader set of characteristics, encompassing teacher knowledge, skill and intelligence, but also including interpersonal relationships with parents, other teachers and the principal, as well as a caring attitude toward students.

The divergence in the factors associated with teacher value added and those which are related to principal evaluation may arise in part because principals consider educational objectives beyond student achievement. We find differences between principals' overall assessments of teachers and principals' impressions of how well the same teachers raise student achievement. The correlation between the two assessments is relatively strong, however, and

varies in predictable ways. Principals appear to better at evaluating teachers in elementary school settings, where there are likely greater opportunities for principals to directly observe the classroom behavior of their teachers. Similarly, there is some evidence that the correlation between principal ratings and teacher value added grows stronger with the years of exposure to specific teachers and the principal's overall experience as an administrator.

The relative importance of intelligence, subject knowledge and teaching skills in determining math teacher productivity has important implications for recruiting and preparing future teachers. 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.

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 evaluating current teachers. Principal evaluations could be an important component of retention and tenure decisions if they either measure the same factors more precisely than do past value-added or if they encompass a broader range of factors that are important determinants of teacher productivity. We find some support for both of these possibilities. When value-added measures are constructed from multiple years of test score data, past value-added does a much better job at predicting future teacher performance than do principal evaluations. However, if one only uses as single year of information to estimate teacher value-added, principal evaluations outperform past value-added in predicting future math teacher productivity. In both math and reading,

principal ratings add information that significantly improves the ability to predict future teacher performance compared to using single-year value-added estimates alone. Thus when evaluating early career teachers, for whom there may be only a year or two of student test-score information, the use of the principal evaluations may be particularly valuable.

While this analysis is informative regarding the various ways that teachers could be assessed, it is important to be cautious in drawing conclusions from these results regarding educational policies. For example, the fact that principals' assessments are positively related to value-added, and are sometimes better predictors of future value-added than other indicators, does not necessarily mean that evaluating teachers based on principals' assessments would be a wise policy. The assessments that principals offered in our study involved no financial or employment implications for teachers and it is likely that the principals' stated judgments would differ in a high-stakes context. Also, even if principals would give the same assessments in high-stakes settings, doing so could influence the relationship between principals and teachers in unproductive ways. Nevertheless, the fact that principal evaluations are better predictors of a teacher's contribution to student achievement than are traditional teacher credentials does not lend much support to current policies that reward teachers based on experience and formal education. The subjective principal ratings and objective value-added measures considered here are therefore worth considering as alternatives to the present system of teacher compensation.

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.

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, Don, Pam Grossman, Hamp Lankford, Susanna Loeb and Jim Wyckoff (2008). "Overview of Measuring Effect Sizes: The Effect of Measurement Error." Policy Brief No. 2. Washington, DC: CALDER.

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 (forthcoming). "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects." *Journal of Human Resources*.

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?" Working Paper 9. Washington, DC: National Center for the Analysis of Longitudinal Data in Education Research,

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.

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., 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, 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. and Tim R. Sass (2006). "Value-Added Models and the Measurement of Teacher Quality." Unpublished. Tallahassee, FL: Florida State University.

Harris, Douglas N. and Tim R. Sass (2008). "Teacher Training, Teacher Quality and Student Achievement." Working Paper No. 3. Washington, DC: Center for the Analysis of Longitudinal Data in Education Research (CALDER).

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., Jonah E. Rockoff and Douglas O. Staiger. 2006. "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." Working Paper #12155.

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." Unpublished manuscript.

Koedel, Cory and Julian Betts (2009). "Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique." Unpublished manuscript.

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.

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 (2010). "Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools." Working Paper #16240. Cambridge, MA: National Bureau of Economic Research.

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.

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 using maximum likelihood method. Factor loadings based on promax rotation.

Table 3
Pairwise Correlation of Estimated Teacher Fixed Effects and Principal's Rating of Teachers With Teacher Characteristic Factors

	Estimated Teacher FE	Overall Rating	Ability to Raise Test Scores	Inter-personal Skills	Moti-vation/ Enthus-iasm	Works Well With Others	Knowledge/ Teaching Skills/ Intelligence
			Math				
Estimated Teacher FE	1.000						
Overall Rating	0.276** [0.319]	1.000					
Ability to Raise Test Scores	0.243** [0.280]	0.733**	1.000				
Interpersonal Skills	0.209** [0.241]	0.703**	0.550**	1.000			
Motivation/Enthusiasm	0.192** [0.221]	0.738**	0.596**	0.734**	1.000		
Works Well With Others	0.205** [0.237]	0.762**	0.598**	0.756**	0.732**	1.000	
Knowledge/Teaching Skills/ Intelligence	0.275** [0.317]	0.881**	0.752**	0.612**	0.682**	0.644**	1.000
			Reading				
Estimated Teacher FE	1.000						
Overall Rating	0.168** [0.236]	1.000					
Ability to Raise Test Scores	0.178** [0.250]	0.741**	1.000				
Interpersonal Skills	0.084 [0.118]	0.709**	0.626**	1.000			
Motivation/Enthusiasm	0.138** [0.195]	0.856**	0.702**	0.631**	1.000		
Works Well With Others	0.089 [0.125]	0.697**	0.569**	0.716**	0.684**	1.000	
Knowledge/Teaching Skills Intelligence	0.020 [0.027]	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 4
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		3.136***			1.570***	
		(4.46)			(3.14)	
Ability to Raise Test Scores			2.778***			1.516**
			(3.73)			(2.90)
1-2 Years of Experience	11.696	14.465	12.620	2.933	2.516	-0.364
	(1.19)	(1.52)	(1.26)	(0.45)	(0.39)	(0.04)
3-5 Years of Experience	10.757	11.264	12.951*	5.348	4.585	2.941
	(1.50)	(1.63)	(1.78)	(1.08)	(0.94)	(0.44)
6-12 Years of Experience	9.948	11.908*	10.721	3.365	3.367	1.446
	(1.43)	(1.78)	(1.53)	(0.69)	(0.71)	(0.22)
13-20 Years of Experience	10.425	12.643*	13.088*	3.776	3.780	1.626
	(1.50)	(1.89)	(1.86)	(0.78)	(0.79)	(0.24)
21-27 Years of Experience	14.256**	15.410**	15.856**	4.747	4.034	2.695
	(2.04)	(2.30)	(2.26)	(0.98)	(0.85)	(0.40)
28+ Years of Experience	11.677	13.742**	12.508*	5.136	4.903	2.298
	(1.64)	(2.01)	(1.73)	(1.04)	(1.01)	(0.34)
Advanced Degree	-1.914	-2.012	-1.472	0.028	-0.081	-0.630
	(1.31)	(1.44)	(1.01)	(0.03)	(0.08)	(0.61)
R-squared	0.037	0.114	0.121	0.014	0.056	0.056
No. of Observations	237	237	202	231	231	201

Note: Absolute values of t-ratios 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 5
Quintile Rankings of Teacher Fixed Effects by Principal Ratings of Teachers

Math					
	Principal's "Ability to Raise Test Scores" Quintile Ranking				
Teacher Fixed Effects Quintile Ranking	Bottom 20%	Second 20%	Third 20%	Fourth 20%	Top 20%
Bottom 20%	**41**	25	8	10	16
Second 20%	43	**27**	14	5	11
Third 20%	41	20	**11**	7	20
Fourth 20%	36	23	20	**14**	7
Top 20%	16	28	23	23	**9**
Reading					
	Principal's "Ability to Raise Test Scores" Quintile Ranking				
Teacher Fixed Effects Quintile Ranking	Bottom 20%	Second 20%	Third 20%	Fourth 20%	Top 20%
Bottom 20%	**33**	30	12	2	23
Second 20%	32	**37**	7	5	20
Third 20%	38	15	**18**	10	20
Fourth 20%	37	32	10	**15**	7
Top 20%	15	33	15	21	**15**

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

Table 6
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		3.947***			1.978***	
		(4.90)			(3.42)	
Overall Rating × Middle/High		0.673			0.397	
		(0.48)			(0.41)	
Ability to Raise Test Scores × Elementary			3.781***			2.169***
			(4.42)			(3.49)
Ability to Raise Test Scores × Middle/High			0.039			-0.026
			(0.03)			(0.03)
1-2 Years of Experience	11.696	14.060	12.685	2.933	3.433	2.293
	(1.19)	(1.49)	(1.28)	(0.45)	(0.53)	(0.26)
3-5 Years of Experience	10.757	10.424	12.707*	5.348	4.920	4.282
	(1.50)	(1.52)	(1.76)	(1.08)	(1.01)	(0.63)
6-12 Years of Experience	9.948	11.772*	10.430	3.365	3.900	3.273
	(1.43)	(1.77)	(1.50)	(0.69)	(0.82)	(0.49)
13-20 Years of Experience	10.425	11.846*	12.335*	3.776	4.151	3.242
	(1.50)	(1.78)	(1.77)	(0.78)	(0.87)	(0.48)
21-27 Years of Experience	14.256**	14.923**	15.261**	4.747	4.661	4.214
	(2.04)	(2.24)	(2.20)	(0.98)	(0.98)	(0.63)
28+ Years of Experience	11.677	13.131*	12.457*	5.136	5.403	4.146
	(1.64)	(1.93)	(1.75)	(1.04)	(1.11)	(0.61)
Advanced Degree	-1.914	-1.903	-1.442	0.028	0.022	-0.389
	(1.31)	(1.37)	(1.00)	(0.03)	(0.02)	(0.38)
R-squared	0.037	0.130	0.144	0.014	0.065	0.074
No. of Observations	237	237	202	231	231	201

Note: Absolute values of t-ratios 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 7
FLGS 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	0.050**		0.045*	0.030		0.010
	(2.30)		(1.85)	(0.97)		(0.31)
Principal's Total Experience in Ed. Administration		0.018	0.006		0.028*	0.025
		(1.31)	(0.43)		(1.87)	(1.53)
R-squared	0.187	0.069	0.192	0.040	0.132	0.132
No. of Observations	25	25	25	25	25	25
Correlation Between Teacher Effects and "Ability to Raise Test Scores" Rating						
Principal's Tenure at School	0.012		-0.009	0.091**		0.057
	(0.44)		(0.28)	(2.61)		(1.38)
Principal's Total Experience in Ed. Administration		0.021	0.024		0.056**	0.037
		(1.30)	(1.27)		(2.67)	(1.46)
R-squared	0.009	0.075	0.081	0.245	0.254	0.319
No. of Observations	23	23	23	23	23	23

Note: Absolute values of t-ratios 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 8
Ordinary Least Squares Estimates of the Determinants of Principal's Overall Rating of Teachers
(Grades 2 – 10, 1999/2000 – 2005/06)

	Math		Reading	
	[1]	[2]	[3]	[4]
Ability to Raise Test Scores	0.733***		0.730***	
	(14.59)		(15.18)	
Ability to Raise Test Scores × Elementary		0.755***		0.771***
		(12.95)		(13.33)
Ability to Raise Test Scores × Middle/High		0.670***		0.637***
		(6.95)		(7.26)
R-squared	0.539	0.540	0.564	0.568
No. of Observations	202	202	201	201

Note: Absolute values of t-ratios 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 9
Ordinary Least Squares Estimates of the Determinants of Principal's Overall Rating of Teachers
(Grades 2 – 10, 1999/2000 – 2005/06)

	Math		Reading	
	[1]	[2]	[1]	[2]
Interpersonal Skill	0.096**		0.187***	
	(2.05)		(3.32)	
Knowledge/Teaching Skills/	0.608***		0.601***	
Intelligence	(15.36)		(12.57)	
Motivation/Enthusiasm	0.054		0.024	
	(1.19)		(0.46)	
Works Well With Others	0.233***		0.156***	
	(4.94)		(2.88)	
Interpersonal Skill		0.108**		0.130**
× Elementary		(2.08)		(2.10)
Interpersonal Skill		0.051		0.505***
× Middle/High		(0.41)		(3.69)
Knowledge/Teaching Skills/		0.615***		0.613***
Intelligence × Elementary		(13.80)		(12.20)
Knowledge/Teaching Skills/		0.599***		0.440***
Intelligence × Middle/High		(6.26)		(3.03)
Motivation/Enthusiasm		0.043		0.056
× Elementary		(0.87)		(0.97)
Motivation/Enthusiasm		0.176		-0.048
× Middle/High		(1.30)		(0.44)
Works Well With Others		0.248***		0.190***
× Elementary		(4.78)		(3.28)
Works Well With Others		0.112		-0.031
× Middle/High		(0.89)		(0.22)
R-squared	0.852	0.854	0.805	0.814
No. of Observations	207	207	203	203

Note: Absolute values of t-ratios 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 10
FGLS Estimates of the Determinants of Teacher Fixed Effects
(Grades 2 – 10, 1999/2000 – 2005/06)

	Math		Reading	
	[1]	[2]	[1]	[2]
Interpersonal Skill	1.034		0.619	
	(0.82)		(0.66)	
Knowledge/Teaching Skills/	2.364**		-1.061	
Intelligence	(2.22)		(1.16)	
Motivation/Enthusiasm	-0.261		1.390*	
	(0.21)		(1.74)	
Works Well With Others	0.263		0.483	
	(0.20)		(0.55)	
Interpersonal Skill		1.203		0.593
× Elementary		(0.86)		(0.56)
Interpersonal Skill		3.008		2.143
× Middle/High		(0.94)		(0.93)
Knowledge/Teaching Skills/		2.959**		-1.018
Intelligence × Elementary		(2.46)		(1.01)
Knowledge/Teaching Skills/		-0.203		-1.496
Intelligence × Middle/High		(0.08)		(0.67)
Motivation/Enthusiasm		0.047		1.732**
× Elementary		(0.03)		(2.02)
Motivation/Enthusiasm		-2.590		-0.690
× Middle/High		(0.74)		(0.30)
Works Well With Others		0.273		0.762
× Elementary		(0.19)		(0.77)
Works Well With Others		0.626		-0.441
× Middle/High		(0.19)		(0.27)
R-squared	0.124	0.146	0.063	0.083
No. of Observations	207	207	203	203

Note: Absolute values of t-ratios 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 11
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	0.166** [0.198]	1.000			
Ability to Raise Test Scores	0.141 [0.168]	0.674**	1.000		
Average DSS Gain (2004/05)	0.294** [0.350]	0.189**	0.108	1.000	
Average DSS Gain (2005/06)	0.239** [0.285]	0.055	0.121	0.370**	1.000
			Reading		
Estimated Teacher FE	1.000				
Overall Rating	0.125 [0.180]	1.000			
Ability to Raise Test Scores	0.154 [0.221]	0.676**	1.000		
Average DSS Gain (2004/05)	0.024 [0.034]	0.146	0.063	1.000	
Average DSS Gain (2005/06)	0.271** [0.390]	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 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 (from 00/01-05/06)	0.710*** (5.81)		0.707*** (5.55)	0.640*** (3.77)		0.576*** (3.28)
Principal's Overall Rating (Summer 2006)		2.406 (1.44)	0.104 (0.07)		3.446** (2.26)	2.154 (1.42)
R-squared	0.167	0.012	0.167	0.088	0.034	0.101
No. of Observations	170	170	170	149	149	149
Prior Value-Added Based on Up to Two Years of Teacher Performance						
Prior Value-Added (from 04/05-05/06)	0.478*** (6.66)		0.469*** (6.46)	0.214** (2.14)		0.180* (1.79)
Principal's Overall Rating (Summer 2006)		2.385 (1.43)	1.139 (0.79)		3.43** (2.28)	3.020* (1.96)
R-squared	0.210	0.012	0.213	0.030	0.035	0.056
No. of Observations	169	169	169	148	148	148
Prior Value-Added Based on One Year of Teacher Performance						
Prior Value-Added (from 05/06)	-0.014 (0.72)		-0.012 (0.64)	0.125*** (3.07)		0.119*** (2.97)
Principal's Overall Rating (Summer 2006)		3.146* (1.85)	3.095* (1.81)		3.894** (2.49)	3.543** (2.38)
R-squared	0.004	0.024	0.027	0.074	0.050	0.118
No. of Observations	140	140	140	120	120	120

Prior Value-Added Based on Four Years of Teacher Performance, Last Year Excluded						
Prior Value-Added (from 01/02-04/05)	0.536*** (4.44)		0.531*** (4.21)	0.590*** (3.98)		0.528*** (3.46)
Principal's Overall Rating (Summer 2006)		2.075 (1.25)	0.261 (0.16)		3.787** (2.46)	2.432 (1.60)
R-squared	0.107	0.009	0.107	0.101	0.041	0.117
No. of Observations	166	166	166	144	144	144

Note: Absolute values of t-ratios 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.