

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: July 17, 2007

Preliminary Draft: Please do not cite without permission.

Abstract: There is considerable interest among researchers and policymakers in identifying effective teachers and their characteristics, which could inform the design of teacher training programs, teacher hiring and evaluation systems, and (optimal) labor contracts. We study these issues using student achievement data from a mid-sized Florida school district to calculate teacher “value-added” and to compare these with subjective principal ratings of teachers identified from confidential interviews. We find a positive and significant correlation between teacher value-added and principals’ subjective ratings and that principals’ evaluations are generally, though not always, better predictors of a teacher value-added than traditional approaches to teacher compensation that focus on experience and formal education. Second, principals give considerable weight in their subjective assessments to teachers’ ability to raise student test scores. Third, principals’ ratings of teachers are most closely associated with a combination of teachers’ subject knowledge, teaching skill, and intelligence and to a lesser extent with teachers’ ability to work with others. Likewise, among various teacher traits, the combination of subject knowledge, teaching skill, and intelligence has the strongest and most consistent association with a teacher’s contribution to student learning. Finally, overall principal ratings of teachers perform as well or better than historic value-added estimates at predicting current teacher value-added. Overall, these results suggest that, when assessing their teachers, principals in strong accountability contexts give considerable weight to student test scores and to the teacher characteristics that are most closely related to a teacher’s contribution to student achievement.

*This study is funded under grant R305M040121 from the U.S. Department of Education. The authors wish to thank Stacey Rutledge and Kyle Ingle for their comments and Micah Sanders, Cynthia Thompson and Lisa Verdon for valuable research assistance. All remaining errors are the responsibility of the authors.

I. Introduction

There is general agreement that teachers play an important role in determining the educational outcomes of students, yet relatively little is known about what characteristics make some teachers more effective in promoting student achievement than others. If attributes that influence teacher quality could be determined, they could be used to identify the most desired candidates in the hiring process and to produce better mechanisms for evaluating and compensating existing teachers. Because some 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.

While the extant literature on teacher quality is voluminous, there is little agreement about the factors that contribute significantly to a teacher's ability to raise 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 found that gains in student achievement are correlated with teacher experience (Rockoff (2004), Clotfelter, Ladd and Vigdor (2005), Jepsen (2005), Rivkin, Hanushek and Kain (2005), Harris and Sass (2007)), but little else in the way of observed teacher characteristics seems to consistently matter. In short, while teachers significantly impact student achievement, the variation in teacher effectiveness is still largely unexplained by commonly measured characteristics.

One possible explanation for the inability of extant research to identify the determinants of teacher quality is that researchers have not been measuring the characteristics that truly determine teacher effectiveness. Previous studies have focused primarily on readily observed

characteristics like experience, educational attainment, certification status and college major. Like previous research, we employ data from administrative records to control for many of these “objective” teacher characteristics. However, we also utilize data from principal interviews to explore a number of more “subjective” teacher attributes, including personality traits, teacher skills, and perceived intelligence.

In addition to potentially uncovering characteristics that are associated with teacher effectiveness, the simultaneous analysis of personnel records and principals’ evaluations of teachers can also serve to inform current policy debates on the optimal role for principals in the evaluation and compensation of teachers. Policymakers and researchers have shown great interest in using measures of student achievement in the process of evaluating and paying teachers (e.g., Gordon, Kane, and Staiger (2006)). However, there are concerns about the precision of these measures, their narrow focus on student test scores, and the fact that they can only be calculated for a small proportion of all teachers. Alternatively, both through the creation of charter schools nationwide and through major school reform in school districts such as New York City, there is movement to grant principals greater authority in hiring, evaluation and retention of teachers. The downside of subjective evaluations by principals is they may be subject to personal bias and some principals may simply be poor judges of teacher effectiveness. To address these issues we analyze the relationship between principal evaluations of teachers and the contribution of teachers to student achievement or teacher “value added.” We also compare the ability of past value-added measures and principal ratings to predict future teacher effectiveness.

We begin by estimating a model of student achievement that includes fixed effects to control for unmeasured student, teacher and school heterogeneity. The resulting estimated

teacher fixed effects are our measure of teacher effectiveness or value added. We then analyze the simple correlation between principals' subjective assessments and teachers' value-added scores. Previous work suggests that there is a positive, but arguably low correlation between the two (Jacob and Lefgren (2005)). This is followed by a multivariate analysis to examine whether principals are better at judging teacher effectiveness than traditional approaches to compensation that focus on experience and formal education. Next we look in detail at specific teacher attributes and how they relate to both principals' overall evaluations of their faculty members as well as to estimates of teacher value added. Finally, we compare the ability of principal ratings and past teacher effectiveness measures to predict student achievement.

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

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 (2005), this research suggests that there is a relatively weak relationship between subjective and objective measures (Bommer (1995), Heneman (1986)). One reason is that evaluators appear to report a narrower range of productivity levels than the objective data suggest, perhaps because the risk-aversion of workers leads employers to inflate subjective measures of productivity for the least effective workers and therefore reduce worker uncertainty about employment and compensation (Levin (2003)),

Macleod (2003)).¹

There is a limited literature that specifically addresses the relationship between subjective and objective assessments of school teachers. Some older studies have examined the relationship between the level of student test scores and principals' subjective assessments (Armor et al. (1976), Murnane (1975)) using cross-sectional data. One problem with cross-sectional analysis of test score levels is that such an approach does not account for the influence of prior educational inputs, particularly past teachers, on current performance. Further, use of cross-sectional data prohibits the use of fixed or random effects estimators to control for unobserved student heterogeneity.

The growth of annual student testing has made it possible to track individual students over multiple years. These panel data provide the opportunity to isolate teacher effectiveness from other time invariant factors such as the unmeasured differences in student and family characteristics. Some more recent studies have utilized data of this type to account for selection bias (Medey and Coker (1987), Peterson (1987, 2000)), but, as noted by Jacob and Lefgren (2006), 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 effectiveness that is linked to principals' subjective assessments. They obtain student achievement data and combine it with data on principals' ratings of 202 teachers in a mid-sized school district in a Western state.²

¹ The comparison of subjective and objective measures has also been used to study whether 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 other than their actual productivity (e.g., Varma and Stroh (2001)).

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

They reach three main conclusions: (1) there is a positive correlation between the subjective and objective measures;³ (2) that this correlation holds even after controlling for teacher experience and education levels (the primary bases for determining teacher compensation); and (3) that the principal's evaluation is better than value-added in predicting parent requests for teachers.

A number of other studies have examined the relationship between teacher value-added and subjective teacher ratings that are based on formal standards and extensive classroom observation (Gallagher, 2004; Kimball et al., 2004; Milanowski, 2004).⁴ 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.

Of the above studies, only Jacob and Lefgren consider specific teacher characteristics, as opposed to the overall rating. Their measures of teacher characteristics include: dedication/work ethic, classroom management, providing a role model for students, positive relationships with teacher colleagues and administrators, and the number of requests for the teacher received from parents. 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.)

The present study also estimates the relationship between subjective ratings and teacher value-added, but we go beyond previous research in a number of ways. First, we consider a

³ Another study that has reached this conclusion is Hoffman and Oreopoulos (2006) who find that college students' probability of dropping a course, or taking courses in the same subject in subsequent years, is unrelated to whether the instructor is part-time or full-time, does research, has tenure, or is highly paid. Not surprisingly, students' subjective assessments of instructors are more closely related to these same outcomes. We do not discuss this study in depth because of its focus on the college level and the fact that the relevant instructor characteristics (e.g., salary and experience) are fairly standard and widely considered in past studies of teachers.

⁴ 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.

broader range of teacher characteristics, one that is based on previous theories and evidence of teacher effectiveness.⁵ We include personality traits, such as “caring,” “enthusiastic,” and “intelligent,” as well as evaluations of subject matter knowledge and teaching skill. Second, we analyze the relationship between each of these measures and both the overall evaluation of principals as well as teacher value added. Third, we analyze teacher ratings and student performance in middle and high school, in addition to elementary school. We allow the relationship between teacher characteristics and teacher ratings or teacher value added to vary across these grade groupings to see if the relative importance of characteristics varies across grade levels. Finally, given the relatively long panel of available data we can control for student heterogeneity via student fixed effects when evaluating the performance of teachers and the subjective evaluations of teachers by principals.

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 data from 30 principals from an anonymous mid-sized Florida school district. As described in Table 1, the district includes a diverse population of students,

⁵ For an extensive review of this literature see Harris, Rutledge, Ingle, and Thompson (2006)

teachers, and principals. 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).⁶ However, the percentage female is somewhat larger (sample district: 63 percent; national: 44 percent).

Characteristics of the principals' schools are also in the table, including the grade levels, Title I status, and school grades assigned through Florida's accountability system. These grades—from a high of “A” to a low of “F”—are based primarily on student scores on math, reading, and writing on the state's standardized test, the Florida Comprehensive Assessments Test (FCAT). Some characteristics of the sample are not mentioned or masked to ensure district anonymity.

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*. 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 test scores

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.

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 current policy considerations than would analyses of districts in relatively low-accountability states.

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 2004/05.⁷ Thus we are able to estimate the determinants of achievement gains for five years, 2000/01-2004/05. Characteristics of the sample used in the value-added analysis are described in Table 2.

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

$$\Delta A_{it} = \beta_1 \mathbf{X}_{it} + \beta_2 \mathbf{P}_{-ijmt} + \gamma_i + \delta_k + \phi_m + \nu_{it} \quad (1)$$

⁷ Prior to 2004/05 version 9 of the Stanford Achievement Test (SAT-9) was administered. In 2004/05 the SAT-10 was given. All SAT-10 scores have been converted to SAT-9 equivalent scores.

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, v_{it} , is a normally distributed, mean zero 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). The model is estimated by demeaning the student effects and including explicit indicator variables for teachers and students. To make the estimation computationally tractable we employ the sparse-matrix Stata procedure developed by Cornelißen (2006).

To gauge the relative importance of characteristics that contribute to a teacher’s impact on student achievement, we conduct a two-step estimation procedure first proposed by Dickens and Ross (1984). The estimated teacher fixed effect or “value added” from the student achievement model is regressed on individual teacher credentials, personality traits, teaching skills and perceived intelligence. Following Dickens and Katz (1986), this second-stage

regression is estimated by weighted least squares, with the square root of the numbers of students per teacher as weights.⁸

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 interviewer never knew the name of the teacher being discussed.⁹

From the administrative data described above, we identified teachers in tested grades and subjects in the 30 schools who had data sufficient to estimate teacher value-added and who were still in the school in the last year for which the administrative data were available, 2004-05. 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.

⁸ This procedure effectively gives less weight to less precise estimates of teacher value added. It does not, however, account for the differences in estimated teacher effects due to estimation error. In future work we plan to adopt the empirical Bayes approach used by Jacob and Lefgren to account for the error.

⁹ The interviewer gave the principal a sealed envelope from the district staff that included a list of non-identifiable, unique identifying numbers and corresponding names. The interviewer list included only the numbers.

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.¹⁰ 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 (2006) to be among the most important characteristics that principals look for when hiring teachers.¹¹

The interview questions were designed so that principals would evaluate teachers relative to others on the list. 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, therefore, a common scale. Like Jacob and Lefgren, we therefore 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,

¹⁰ 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."

¹¹ As described in Harris, Rutledge, Ingle and Thompson (2006), the data in this study came from the second in a series 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 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.

normalizing the ratings allow us to compare within-school ratings to within-school teacher value-added.¹²

The final activity of the interview involved asking the principals to rate each teacher according to 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 are intended to help us replicate the Jacob and Lefgren analysis and provide a basis for determining 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 students test scores.

Finally, as part of the interview, we discovered that principals have access to a district-purchased software program, SnapshotTM, 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 at the achievement gains made by students of each teacher. This likely influenced their responses to some of the interview questions. We discuss some possible instances below and plan further analysis to address this.

¹² An alternate approach would be to compare the raw teacher ratings with teacher fixed effects estimated from an achievement model that excludes school fixed effects. This alternative requires one to assume that principals use a common rating metric.

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 2004/05. Estimates of this first-stage model are presented in Table 3. For both math and reading, student fixed effects account for about 18 percent of the total variation in test score gains. In contrast, about 10 percent of the variance in reading gains and 3 percent of the variance in math gains are accounted for by the teacher fixed effects.

In order to reduce potential multicollinearity problems and reduce the number of teacher characteristics to analyze, we follow Jacob and Lefgren and conduct a factor analysis of the 11 individual teacher characteristics rated by principals. As indicated in Table 4, the individual characteristics can be summarized into four factors: interpersonal skills, motivation/enthusiasm, ability to work with others, and knowledge/teaching skills.

Simple correlations between the estimated teacher fixed effects, principals' overall ratings of teacher quality, 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 5. Looking at the first column, it appears that 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. The correlation between value-added and the principals impression of the teacher's ability to raise test scores is similar. Part of the explanation is probably that simple mean gains lead to quite different measures of teacher contributions to test scores than do teacher fixed effects. It could also mean that principals are not using the software program to determine by-teacher average student achievement.

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 weight 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. It is possible there is 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 of all the four factors, knowledge/teaching skills is most closely associated with teacher value-added. This result holds up in the regression analyses below.

Table 6 presents estimates of the determinants of the teacher fixed effects, including only standard teacher credentials (experience, possession of an advanced degree, certification status) along with general principal evaluations (overall rating, ability to raise test scores) as explanatory variables. The first column reports estimates where only teacher credentials and no principal ratings are included. None of the variables are statistically significant determinants of teacher value added scores.¹³ In contrast, when a principal's overall rating of a teacher is added to the model, its coefficient is positive and highly significant in both reading and math. This suggests that principals have knowledge about teacher quality that is not captured by the standard

¹³ In another study using statewide data from Florida we find that the effects of teacher experience are highly significant when teacher fixed effects are excluded, but within-teacher changes in experience are often not statistically significant (Harris and Sass, 2007). 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 picking up idiosyncratic features of the handful of reading teachers in the sample who are not fully certified during part of the sample period.

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

As expected, the principal's perception of a teacher's ability to raise math test scores is positively and significantly correlated with the value added of math teachers. This is not the case in reading, however, where the point estimate is smaller and statistically insignificant. One possible explanation for this surprising result is that it is more difficult to determine which teachers are responsible for the school's contribution to reading, especially in middle and high school where there are few classes that are clearly labeled as "reading." Evidence in Table 7 below reinforces this interpretation.¹⁴

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 6 is +2.685, which implies that a teacher who is rated one point higher on the 1-9 scale raises student math test scores by 2.685 scale score points per year more than the average teacher, which translates to 0.05 standard deviations.¹⁵ While this might be considered small by some standards, these represent only single-year changes, which could accumulate to relatively larger effects over time.

In Table 7 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, all four principal ratings (overall and ability to raise test scores in reading and math) are positively and statistically significantly associated with the teacher

¹⁴ A second possible issue is that estimating the school's contribution to reading is more difficult because reading scores are widely believed to be influenced by factors outside of school. Students may read books in their free time, but they are unlikely to do math problems. However, this should be captured by the student fixed effects.

¹⁵ This conversion is based on the standard deviation in the level of math achievement, 55.72. The standard deviation in the level of reading achievement is 51.34. One might also be interpreted in interpreting these in terms of the standard deviations in the gains which are 32.40 for reading and 30.53 for math.

fixed effect. For middle and high school teachers, there are no significant relationships. This difference across grade levels is partly due to the smaller number of teachers. Eighteen of the 30 schools are elementary and each school has roughly the same number of teachers in the analysis. But this is not the only explanation as all the point estimates are lower in reading. This is especially true in the upper grades, as expected given the greater difficulty of identifying which teacher is responsible for reading in those grades. The smaller point estimates in reading compared with math in the elementary grades, where matching teachers with subjects is much easier, suggests a possible difference in the scale of the tests or a genuine difference in the relationship between value-added and principal ratings.

We next turn to an analysis of the factors affecting a principal's overall rating of a teacher. Table 8 presents least squares estimates of the relationship between perceived ability to raise test scores in the relevant subject and the principal's overall rating of teacher quality. 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 55 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. As noted in Table 5, the four teacher attributes are highly correlated, which will introduce the problem of multicollinearity. In both subjects, knowledge/teaching Skills contributes the most to the principals' overall rating, followed by works well with others. There are some apparent differences by grade level, though none of these differences is statistically significant. Also, note that the explained variation of roughly 0.8

reinforces the validity of the 12 characteristics as the factors that determine principals' overall ratings.

Very different patterns emerge when we switch the dependent variable to the teacher fixed effect in Table 10. Column [1] suggests that knowledge/teaching Skills is positively and significantly associated with teacher value-added in reading. None of the other coefficients in column [1] are significant. Column [2] shows that the effect of knowledge/teaching skills in reading occurs mainly in the elementary grades. Some of the coefficients that were insignificant in math when the grade levels were pooled become significant when the effects are separated by grade level. Interestingly, teachers with high ratings in motivation/enthusiasm have much lower math value-added, while teachers who work well with others have much higher value-added. The overall explanatory power of the four factors is quite low, however.

In Table 11 we examine the relationship between teacher value-added and each teacher characteristic factor separately. This has little effect on the reading results, but changes the math results dramatically. All four coefficients are now positive and significant in the math columns. This may explain why previous studies have found inconsistent relationships between teacher characteristics and teacher value-added. By omitting one or more factors, the coefficients on the included variables appear to become more positive and significant. This suggests that including all four factors simultaneously may have lead to multicollinearity.

Finally, in Table 12 we compare the ability of principal ratings and past value added measures to predict future student achievement. In each case we estimate a model of student achievement gains in 2004/05, including either the principal's overall rating of the teacher or the estimated teacher fixed effect from a student achievement model covering the years 1999/00-2003/04. Although the model is estimated for a single year of achievement gains, teacher fixed

effects from the prior 1999/2000-2003/04 period are included to control for unobserved student heterogeneity. Both principal ratings and prior teacher value added scores are found to be significant determinants of future value added. The fit of each model is nearly identical, explaining about 21 percent of the variation in student achievement. Lagged value-added is insignificantly related to future value-added in reading, while the coefficient in math is significant but apparently small in magnitude. This appears to reflect some instability in teacher value-added over time.

V. Summary and Conclusions

In this paper we analyze the relationship between principals' ratings of teachers and the estimated contribution of teachers to student achievement or "value added" based on standardized tests. As in most previous studies of the subject, we find a positive and significant relationship between the two types of measures. We also find that principals' assessments are better predictors of teacher value-added than are teachers' levels of education and experience and, in some cases, a better predictor than lagged value-added.

The relationship between the subjective and objective measures is still far from perfect, however, and we explore a variety of reasons why they differ. First, principals consider educational objectives beyond student achievement, as reflected by the differences between principals' overall assessments of teachers and principals' impressions of how well the same teachers raise student achievement. The correlation is relatively strong, however, and this may reflect both a desire of principals to be consistent in their various ratings of individual teachers and the incentives principals face under Florida's test-based accountability system.

Principals also appear to have some difficulty determining how much teachers contribute to student learning in reading, as reflected by the weak and statistically insignificant correlation, between teacher value-added and the principals' assessment of how well teachers raise achievement scores. Some reasons for this include the fact that there are no specific reading courses in middle and high school and that teacher value-added is subject to measurement error that attenuates the correlations. While we plan to explore this in future analysis, evidence from Jacob and Lefgren suggests that measurement error is probably modest.

We also explore the relationship between the various measures of teacher effectiveness and measures of specific teacher characteristics that are not typically available. We find that principals' overall assessments are best predicted by their ratings of teacher knowledge/teaching skills, followed by their ability to work well with others. Only the knowledge/teaching-skills rating is related to teacher value-added and then only in reading. Omitting individual teacher characteristics also appears to make factors such as knowledge/teaching Skills more closely related to value-added. This may explain the inconsistent results in previous studies that have varied in their measured teacher characteristics.

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 apparently better predictors of value-added than many other indicators, does not necessarily mean that rewarding teachers based on principals' assessments would be a wise policy. The assessments that principals offered in the interviews had no financial or employment implications 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 it appears quite difficult to identify characteristics of teachers that are systematically related to value-added does not lend much support to current policies that rely 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.
- 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.
- Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2005. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." Unpublished. Durham: Duke University.
- Cornelißen, Thomas (2006). "Using Stata for a Memory Saving Fixed Effects Estimation of the Three-Way Error Component Model." Unpublished. Hannover: University of Hannover.
- Dickens, William T. and Brian A. Ross (1984). "Consistent Estimation Using Data From More Than One Sample." Technical Working Paper #33. Cambridge, MA: National Bureau of Economic Research.
- Dickens, William T. and Lawrence F. Katz (1986). "Interindustry Wage Differences and Industry Characteristics." Working Paper #2014. Cambridge, MA: National Bureau of Economic Research.
- Gallagher, H. A. (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.
- 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.
- 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 (2006). "Mix and match: What principals look for when hiring teachers," paper presented at the 2006 conference of the American Education Research Association.

- 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 (2007). "Teacher Training, Teacher Quality and Student Achievement." Unpublished. Tallahassee, FL: Florida State University.
- Heneman, Robert L. (1986). "The Relationship Between Supervisory Ratings and Results-Oriented Measures Performance: a Meta-analysis." *Personnel Psychology* 39: 811-826.
- Hoffman, F. and Philip Oreopoulos (2006). "Professor Qualities and Student Achievement." *NBER Working Paper 12596*. Cambridge, MA: National Bureau of Economic Research.
- 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.
- Jepsen, Christopher (2005). "Teacher Characteristics and Student Achievement: Evidence from Teacher Surveys." *Journal of Urban Economics* 57(2):302-19.
- Levin, Jonathan (2003). "Relational incentive contracts." *American Economic Review* 93:835-57.
- MacLeod, W. Bentley (2003). "Optimal Contracting with Subjective Evaluation." *American Economic Review* 93(1): 216-240.
- Medley, Donald M. and Homer Coker (1987). "The Accuracy of Principals' Judgments of Teacher Performance." *Journal of Educational Research* 80(4): 242-247.
- Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student assessment: Evidence from Cincinnati. *Peabody Journal of Education*, 79(4), 33-53.
- 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.
- 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.
- Peterson, Kenneth D. (2000). *Teacher Evaluation: A Comprehensive Guide to New Directions and Practices* (2d ed.). Thousand Oaks, CA: Corwin Press.
- Peterson, Kenneth D. (1987). "Teacher Evaluation with Multiple and Variable Lines of Evidence." *American Educational Research Journal* 24(2): 311-317.

- 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.
- 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 School and Principal Characteristics

<i>Level</i>	<i>School Characteristics</i>				<i>Principal Characteristics</i>		
	<i>% Minority</i>	<i>% Free/Reduced Lunch</i>	<i>Title I Status</i>	<i>School Grade</i>	<i>Race</i>	<i>Gender</i>	<i>Years as Principal</i>
Elementary	75	65	Y	B	W	M	14
Elementary	75	70	Y	A	W	F	6
Elementary	30	30	N	A	W	F	8.5
Elementary	44	35	N	A	W	F	5
Elementary	25	20	N	A	W	F	6
Elementary	85	90	Y	A	W	F	4
Elementary	40	70	Y	B	W	M	7
High	15	5	N	A	W	M	4
Middle	70	60	N	A	B	F	6
Elementary	30	15	N	A	W	M	2
Elementary	85	70	Y	B	W	F	1
Elementary	25	10	N	A	W	F	11
Elementary	90	90	Y	C	B	F	4
High	40	15	N	B	W	F	10
K-8	10	55	N	B	W	M	13
Middle	75	55	Y	A	W	M	6
Elementary†	95	90	Y	C	B	F	2
Elementary	50	40	N	A	B	F	3.5
Elementary	95	85	Y	A	B	F	9
Middle	45	35	N	A	W	F	1.5
Middle	30	20	N	A	W	F	11
Middle‡	20	10	N	A	W	M	1
High‡	80	45	N	C	W	M	2
Middle	80	80	Y	C	B	M	1
Elementary	15	5	N	A	W	F	10
Middle	85	80	Y	D	B	F	4
Elementary	50	50	N	A	W	M	12
Elementary	50	45	N	A	W	F	15
Middle	55	40	N	A	W	F	1
High	30	15	N	A	W	M	1
Elementary=17	Mean=53.1	Mean=46.5	Yes=10	A=20	White= 23	Male=11	Mean=6
K-8=1			No=20	B=5	Black=7	Female=19	
Middle=8				C=4			
High=4				D=1			

Notes: Race and free and reduced price lunch percentages rounded to the nearest five to help maintain school and district confidentiality. Data, unless otherwise noted, are from the 2005-2006 academic year.

†Characteristics reflect 2004-2005 data due to school closure. Principal remained in district and was able to evaluate teachers.

‡Assistant principals were interviewed due to time constraints of principals. Principals attested that the interviewed assistant principals took part in annual performance evaluations and had adequate knowledge of sampled teachers.

Table 2
Sample Student and Teacher Characteristics

	Math Sample		Reading Sample	
	No. of Obs.	Mean	No. of Obs.	Mean
Students				
Black	76308	0.352	70916	0.343
Hispanic	76308	0.020	70916	0.020
Free/Reduced Price Lunch	76308	0.312	70916	0.308
Achievement Gain	76308	18.713	70916	15.778
Teachers				
Male	943	0.118	960	0.081
White	943	0.688	960	0.717
Hold Advanced Degree	926	0.348	946	0.370
Years of Experience	943	10.195	960	10.257
Fully Certified	933	0.924	955	0.935
Taught Primarily Elementary School	943	0.718	960	0.706
Taught Primarily Middle School	943	0.152	960	0.153
Taught Primarily High School	943	0.131	960	0.140
Principal's Overall Rating	234	7.103	231	7.108
Rating of Ability to Raise Test Scores	207	7.232	201	7.164
Rating on "Caring"	234	7.397	231	7.468
Rating on "Enthusiastic"	234	7.269	231	7.385
Rating on "Motivated"	234	7.436	231	7.494
Rating on "Strong Teaching Skills"	234	7.560	231	7.597
Rating on "Knows Subject"	234	7.868	231	7.896
Rating on "Communication Skills"	234	7.594	231	7.710
Rating on "Intelligence"	234	7.897	231	7.922
Rating on "Positive Relationship with Parents"	233	7.511	230	7.596
Rating on "Positive Relationship with Students"	233	7.670	230	7.730

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

Table 3
Value-Added Results Used for Estimation of Teacher Effects
(Grades 2 – 10, 1999/2000 – 2004/05)

	Math	Reading
Number of Schools Attended	-2.131** (2.02)	-0.977 (0.85)
Attended Different School in Prior Year	1.910*** (2.99)	0.834 (1.18)
Class Size	-0.111** (2.21)	-0.0380 (0.68)
Proportion of Classroom Peers Who are White	-7.788*** (4.37)	-2.481 (1.30)
Grade-by-Year Indicators	Yes	Yes
Student Fixed Effects	Yes	Yes
Teacher Fixed Effects	Yes	Yes
School Fixed Effects	Yes	Yes
Covariance of Achievement Gain and:		
Student FE	0.185	0.189
Teacher FE	0.031	0.099
Model (Including School Indicators)	0.195	0.165
Error	0.589	0.547
No. of Observations	76,308	70,916

Note: Absolute values of t-ratios adjusted for clustering at the classroom level 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 grade repeaters by grade.

Table 4
Factor Loadings of Normalized Principal Ratings

Teacher Characteristic Rated by Principal	Interpersonal Skills	Motivation/ Enthusiasm	Works Well With Others	Knowledge/ Teaching Skills
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	0.0030	0.0055	0.0332	0.7182
Works Well With Grade Team/Dept.	0.0243	-0.0598	0.8522	0.0838
Works Well With Me (Principal)	0.1390	0.0461	0.8329	-0.0597
Positive Relationship With Parents	0.7554	0.0528	0.0649	0.0757
Positive Relationship With Students	0.9171	0.0271	0.0193	0.0313
Caring	0.6046	0.0983	0.2541	-0.0367
Enthusiastic	0.0775	0.9879	-0.0464	-0.0191
Motivated	0.0247	0.5366	0.2017	0.2445
Strong Teaching Skills	0.2237	0.0174	-0.0802	0.8140
Knows Subject	-0.0834	-0.0156	0.0385	0.9819
Communication Skills	0.1650	0.2167	0.1715	0.3290

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 5
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	Know- ledge/ Teaching Skills
Math							
Estimated Teacher FE	1.000						
Overall Rating	0.203**	1.000					
Ability to Raise Test Scores	0.225**	0.745**	1.000				
Interpersonal Skills	0.165**	0.703**	0.550**	1.000			
Motivation/Enthusiasm	0.159**	0.738**	0.596**	0.734**	1.000		
Works Well With Others	0.182**	0.762**	0.598**	0.756**	0.732**	1.000	
Knowledge/Teaching Skills	0.212**	0.881**	0.752**	0.612**	0.682**	0.644**	1.000
Reading							
Estimated Teacher FE	1.000						
Overall Rating	0.095	1.000					
Ability to Raise Test Scores	0.109	0.736**	1.000				
Interpersonal Skills	0.041	0.719**	0.626**	1.000			
Motivation/Enthusiasm	0.014	0.699**	0.569**	0.714**	1.000		
Works Well With Others	0.004	0.725**	0.589**	0.762**	0.677**	1.000	
Knowledge/Teaching Skills	0.131	0.861**	0.697**	0.644**	0.682**	0.650**	1.000

Note: **indicates significance at the .05 level.

Table 6
Weighted Least Squares Estimates of the Determinants of Teacher Fixed Effects
(Grades 2 – 10, 1999/2000 – 2004/05)

	Math			Reading		
	[1]	[2]	[3]	[1]	[2]	[3]
Overall Rating		2.685*** (2.82)			1.661* (1.76)	
Ability to Raise Test Scores			2.570*** (4.52)			0.975 (1.48)
1-2 Years of Experience	3.770 (0.36)	4.706 (0.46)	2.789 (0.27)	1.926 (0.18)	0.822 (0.08)	-7.885 (0.59)
3-5 Years of Experience	9.101 (1.30)	9.428 (1.36)	9.838 (1.44)	10.257 (1.37)	9.530 (1.27)	5.939 (0.65)
6-12 Years of Experience	3.919 (0.57)	5.092 (0.75)	3.710 (0.54)	8.089 (1.10)	7.815 (1.07)	3.050 (0.34)
13-20 Years of Experience	8.313 (1.24)	9.616 (1.45)	10.503 (1.60)	11.637 (1.60)	11.352 (1.56)	5.375 (0.59)
21-27 Years of Experience	10.124 (1.49)	10.382 (1.55)	10.689 (1.61)	5.116 (0.70)	4.115 (0.57)	0.159 (0.02)
28+ Years of Experience	2.113 (0.30)	3.421 (0.49)	3.736 (0.53)	7.797 (1.04)	7.302 (0.98)	0.045 (0.00)
Advanced Degree	-1.941 (0.99)	-2.055 (1.06)	-1.568 (0.80)	-0.093 (0.05)	-0.306 (0.16)	-0.639 (0.30)
Full Certification	2.229 (0.24)	-0.278 (0.03)	-1.339 (0.15)	-388.413 (1.10)	-417.302 (1.16)	-500.873 (1.35)
R-squared	0.045	0.078	0.149	0.047	0.061	0.057
No. of Observations	234	234	207	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
Weighted Least Squares Estimates of the Determinants of Teacher Fixed Effects
(Grades 2 – 10, 1999/2000 – 2004/05)

	Math			Reading		
	[1]	[2]	[3]	[1]	[2]	[3]
Overall Rating × Elementary		3.328*** (2.84)			2.483** (2.15)	
Overall Rating × Middle/High		1.424 (0.87)			0.029 (0.02)	
Ability to Raise Test Scores × Elementary			4.071** (3.31)			2.971** (2.25)
Ability to Raise Test Scores × Middle/High			1.456 (0.88)			-0.554 (0.32)
1-2 Years of Experience	3.770 (0.36)	3.969 (0.39)	-2.179 (0.20)	1.926 (0.18)	0.853 (0.08)	-7.131 (0.53)
3-5 Years of Experience	9.101 (1.30)	8.632 (1.24)	9.671 (1.35)	10.257 (1.37)	9.113 (1.22)	6.205 (0.68)
6-12 Years of Experience	3.919 (0.57)	4.688 (0.68)	2.789 (0.40)	8.089 (1.10)	7.887 (1.08)	3.804 (0.42)
13-20 Years of Experience	8.313 (1.24)	8.793 (1.32)	8.385 (1.23)	11.637 (1.60)	11.002 (1.52)	6.368 (0.71)
21-27 Years of Experience	10.124 (1.49)	9.878 (1.47)	9.986 (1.46)	5.116 (0.70)	4.249 (0.59)	0.677 (0.08)
28+ Years of Experience	2.113 (0.30)	2.600 (0.37)	1.581 (0.22)	7.797 (1.04)	7.149 (0.96)	1.631 (0.18)
Advanced Degree	-1.941 (0.99)	-1.999 (1.03)	-0.614 (0.31)	-0.093 (0.05)	-0.090 (0.05)	-0.084 (0.04)
Full Certification	2.229 (0.24)	0.964 (0.10)	-2.104 (0.23)	-388.413 (1.10)	-419.225 (1.20)	-463.246 (1.24)
R-squared	0.045	0.082	0.126	0.047	0.067	0.071
No. of Observations	234	234	199	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 8
Weighted Least Squares Estimates of the Determinants of Principal's Overall Rating of Teachers
(Grades 2 – 10, 1999/2000 – 2004/05)

	Math		Reading	
	[1]	[2]	[3]	[4]
Ability to Raise Test Scores	0.369*** (11.48)		0.415*** (12.31)	
Ability to Raise Test Scores × Elementary		0.369*** (11.20)		0.413*** (12.12)
Ability to Raise Test Scores × Middle/High		0.369*** (11.20)		0.421*** (11.93)
R-squared	0.425	0.425	0.468	0.469
No. of Observations	207	207	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 and certification status and a constant term.

Table 9
Weighted Least Squares Estimates of the Determinants of Principal's Overall Rating of Teachers
(Grades 2 – 10, 1999/2000 – 2004/05)

	Math		Reading	
	[1]	[2]	[1]	[2]
Interpersonal Skill	0.091* (1.86)		0.192*** (3.26)	
Knowledge/Teaching Skills	0.592*** (14.53)		0.592*** (11.49)	
Motivation/Enthusiasm	0.069 (1.45)		-0.002 (0.04)	
Works Well With Others	0.237*** (4.83)		0.193*** (3.30)	
Interpersonal Skill × Elementary		0.095* (1.69)		0.116* (1.67)
Interpersonal Skill × Middle/High		0.091 (0.86)		0.433*** (3.59)
Knowledge/Teaching Skills ×Elementary		0.602*** (12.46)		0.599*** (10.54)
Knowledge/Teaching Skills × Middle/High		0.559*** (6.90)		0.474 (3.80)
Motivation/Enthusiasm × Elementary		0.049 (0.91)		0.059 (0.92)
Motivation/Enthusiasm × Middle/High		0.179 (1.58)		-0.074 (0.82)
Works Well With Others × Elementary		0.261*** (4.57)		0.218*** (3.37)
Works Well With Others × Middle/High		0.127 (1.26)		0.067 (0.52)
R-squared	0.853	0.854	0.786	0.794
No. of Observations	207	207	206	206

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 and certification status and a constant term.

Table 10
Weighted Least Squares Estimates of the Determinants of Teacher Fixed Effects
(Grades 2 – 10, 1999/2000 – 2004/05)

	Math		Reading	
	[1]	[2]	[1]	[2]
Interpersonal Skill	-0.250 (0.15)		1.656 (0.93)	
Knowledge/Teaching Skills	1.579 (1.10)		3.044* (1.96)	
Motivation/Enthusiasm	-1.030 (0.61)		-1.532 (0.96)	
Works Well With Others	2.523 (1.46)		-1.745 (0.99)	
Interpersonal Skill × Elementary		-0.944 (0.49)		1.425 (0.67)
Interpersonal Skill × Middle/High		3.508 (0.97)		3.446 (0.94)
Knowledge/Teaching Skills × Elementary		1.324 (0.80)		3.151* (1.83)
Knowledge/Teaching Skills × Middle/High		2.061 (0.74)		3.808 (1.00)
Motivation/Enthusiasm × Elementary		1.408 (0.75)		-0.635 (0.33)
Motivation/Enthusiasm × Middle/High		-11.360*** (2.91)		-3.838 (1.41)
Works Well With Others × Elementary		2.133 (1.09)		-1.129 (0.57)
Works Well With Others × Middle/High		7.260** (2.09)		-5.242 (1.33)
R-squared	0.078	0.125	0.056	0.078
No. of Observations	207	207	206	206

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 and certification status and a constant term.

Table 11
Weighted Least Squares Estimates of the Determinants of Teacher Fixed Effects
(Grades 2 – 10, 1999/2000 – 2004/05)

	Math				Reading			
	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]
Interpersonal Skill	1.866* (1.94)				1.167 (1.15)			
Knowledge/Teaching Skills		2.295** (2.38)				1.881* (1.84)		
Motivation/Enthusiasm			1.661* (1.77)				0.559 (0.56)	
Works Well With Others				2.572*** (2.63)				0.521 (0.50)
R-squared	0.058	0.066	0.055	0.072	0.035	0.045	0.030	0.030
No. of Observations	207	207	207	207	206	206	206	206

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 and certification status and a constant term.

Table 12
Estimates of the Determinants of Student Achievement Gains
(Grades 2 – 10, 2004/05)

	Math		Reading	
	[1]	[2]	[1]	[2]
Overall Rating	2.107*** (3.86)		1.446*** (2.69)	
Teacher Fixed Effect (from 99/00-03/04)		0.230*** (3.85)		0.106 (1.44)
R-squared	0.236	0.237	0.115	0.114
No. of Observations	5361	5361	4399	4399

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 individual student mobility, class size, peer characteristics, student fixed effects (from 1999/00-2003/04), school indicators and a constant term.